Optimizing Computational Frameworks to Study the Influence of the Protein Environment on the Individual Site Energies of Chromophores in Photosystem II of Photosynthesis

by

Andrew Cheesman

BSc. (Honours Physics & Mathematics) Department of Physics,
Brock University, Canada

BSc. (Honours Computer Science) Department of Computer Science,
Brock University, Canada

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in

The Faculty of Mathematics and Science
Department of Physics

BROCK UNIVERSITY
January 21, 2016

Abstract

Photosynthesis is a process in which electromagnetic radiation is converted into chemical energy. Photosystems capture photons with chromophores and transfer their energy to reaction centers using chromophores as a medium. In the reaction center, the excitation energy is used to perform chemical reactions. Knowledge of chromophore site energies is crucial to the understanding of excitation energy transfer pathways in photosystems and the ability to compute the site energies in a fast and accurate manner is mandatory for investigating how protein dynamics effect the site energies and ultimately energy pathways with time. In this work we developed two software frameworks designed to optimize the calculations of chromophore site energies within a protein environment. The first is for performing quantum mechanical energy optimizations on molecules and the second is for computing site energies of chromophores in a fast and accurate manner using the polarizability embedding method. The two frameworks allow for the fast and accurate calculation of chromophore site energies within proteins, ultimately allowing for the effect of protein dynamics on energy pathways to be studied. We use these frameworks to compute the site energies of the eight chromophores in the reaction center of photosystem II (PSII) using a 1.9 Å resolution x-ray structure of photosystem II. We compare our results to conflicting experimental data obtained from both isolated intact PSII core

preparations and the minimal reaction center preparation of PSII, and find our work more supportive of the former.

Contents

List of Tables

List of Figures

1. Introduction

Photosynthesis is an essential process to all life on earth. It is the process by which plants, algae and cyanobacteria harvest solar energy and convert it into chemical energy stored in the form of NADPH (nitcotinamide adenine dinucleotide phosphate) and ATP (adenosine triphosphate), the "currency of energy" in biological systems. The process begins with the absorption of a photon via a chromophore, followed by excitation energy transfer (EET) to a reaction center (RC) where energy conversion occurs [1]. It is a very efficient process with yields greater than 95% for the conversion of the excitation energy into charge separation.

Despite several decades of study, many of the atomistic details of photosynthesis remain elusive. These details include the processes by which absorbed light energy is transferred to RCs. It is known that the energy is transferred through the arrays of many thousands of chromophore molecules and occurs with near unity efficiency in the presence of fluctuating dissipative environment, yet the details of this process are not fully understood.

Chromophores of these systems have a relatively large range of site energies (difference between energy levels of first excited state and ground state) even though there are only a few different types of chromophores (Chlorophyll-a, chlorophyll-b, pheophytin). Variations in chromophore site energies arise from interactions with the local environment. Expanding the range of site energies poses some advantages. Firstly by having a wider range of site energies, photosynthetic systems are capable of absorbing photons from a larger range of the solar spectrum and hence capturing more energy. Secondly by taking advantage of the property in which excitation energy will take preference in transferring to a chromophore with a lower site energy

than its current location (section 2.3.1). With this downhill energy transfer preference, it is possible for systems to influence the EET into favouring pathways which lead the excitation energy to the RC via influencing the relative site energies of the system's chromophores. It is known that some complexes have arranged their chromophores in such a way, as is the case with phycobilisomes, an antenna complex associated with photosystem II (PSII). It is not yet known whether the chromophores in PSII are configured in such a fashion and, this is one of the questions we hope to elucidate.

Detailed modeling of photosynthetic light energy conversion requires knowledge of the excited state energy levels of all chromophores in the smallest photosynthetic functional units, photosystems I (PSI) and II (PSII). There are a large number of pigments in these units (~48 in PSII and ~118 in PSI) [2], [3] with relatively small (but functionally important) differences in excited state energies. This complicates the interpretation of experimental observations and makes spectroscopic assignment of the energy levels of individual pigments unreliable. In addition, optical spectroscopy is not capable of relating energetic differences to molecular features, and thus cannot explain origins of the differences in excitation energies.

Computation of accurate site energies based on X-ray crystallographic structures is currently the best approach to assigning individual site energies to specific chromophores. The major difficulty in a computational analysis of systems such as PSI or PSII (~57500 atoms) arises from limitations of computing power. With today's technology it is not feasible to run quantum mechanical (QM) computations on an entire system of this scale.

As chromophores in PSI and PSII are chemically identical chlorins, differences in their excited state energy levels arise solely from their interactions with the protein they are embedded in. Methods combining a QM treatment of the chromophore with a classical electrostatic treatment of its environment have been used to accelerate computation of site energies in PSI and PSII [4]. However, the application of such methods has been problematic. Up to date computations of excited state energies in PSI and PSII have represented the protein environment as a set of static point charges. This is a simplistic representation of the electric field of a protein and its interaction with electronic transitions. It has been recently shown that both static and dynamic contributions of electronic transitions to the reaction field are required for accurate computation of site energies [5]. Another critical factor, not considered before, is accurate determination of the equilibrated ground state conformations of individual chromophores within the protein environment.

Development of a fast and faithful method for computation of site energies would be beneficial as it would allow for the accurate computation of energy levels while illustrating the contributing molecular factors to these energy levels. Additionally changes to the protein structure could be made easily, allowing investigation into what features of the protein environment are contributing to energy level changes. This would allow comprehension into how photosystems funnel the excitation energy into their reaction centers providing insight and guidance into biotechnology and bioengineering of photosynthetic proteins [6], [7].

The goal of this thesis is to develop a framework that will allow for reliable site energy computations to be performed on large pigment-protein systems. To achieve this, the framework will break the larger systems down into smaller QM clusters to facilitate determination of equilibrated ground state chromophore conformations in

a reasonable time. Then we implement a polarizable embedding model [8] which accurately models the effects of the environment surrounding a chromophore by including them directly in the wavefunction of the chromophore to quantify effects of protein polarization and higher order electric multipoles on site energies. We will then apply our framework to computation of site energies of 8 PSII reaction center chromophores and compare our results with experimental studies.

Research in this direction is expected to elucidate the natural design of the biological energy conversion process, and could be beneficial for the discovery of more efficient photovoltaic materials.

2. Literature Review

2.1. Photosynthetic Chromophores

Chromophores are molecules (or parts of molecules) which absorb photons within the visible spectrum, making them a crucial component of photosynthesis. Chromophores are responsible for absorbing the photons whose energy is used in the photosynthetic process. They are bound to specific proteins which provide scaffolding and ordering of the chromophores in the photosynthetic apparatus. Chromophores are also directly involved in energy transfer from the initially excited chromophore through an array of energetically coupled chromophores within the photosystems to the photoactive reaction center where photochemistry occurs [9]. Energy transfer in photosynthesis has a very high quantum efficiency [10], [11]. The photosynthetic apparatus of higher plants includes light harvesting complexes and two photosystems, photosystem I (PSI) and II (PSII) [12]. There are several different kinds of chomophores found in photosynthetic systems: carotenoids, chlorophylls and pheophytins [13]. Of these chromophores chlorophylls are the principal light-harvesting molecules and the most common in both PSI and PSII. Pheophytins are free base chlorophylls found in the reaction center of PSII. PSII contains 35 chlorophyll, 2 pheophytin and 11 beta-carotene molecules [14] while PSI contains 96 chlorophyll, 22 beta-carotene and no pheophytin [15].

2.2. Chlorophylls

2.2.1. Chlorophyll Structure

Chlorophyll is composed of three parts, a phytyl chain or "tail", a chlorin ring or "head" and side chains connected to the "head" [16]. The structure of chlorophyll can be seen in Figure 2-1. The chlorin ring has a metal ion Mg^{2+} at its center. There are two types of chlorophyll in plants, chlorophyll a (Chl a) and chlorophyll b (Chl b) which are nearly identical with the exception of one of their side chains attached to their chlorin rings, where in Chl a the side chain is -CH_3, while the respective side chain in Chl b is -CHO [16]. Energies of absorbed photons vary for the different types of chlorophyll.

Chl b is not found in either PSI or PSII, but is found in light harvesting antennae complexes around them. As we will be looking specifically at PSII, any reference to Chl from here on in will be in reference to Chl a. It is the chlorin ring of Chl which absorbs the photons of visible light. The Chl "tail" is hydrophobic and assists in keeping the Chl in place which it achieves by anchoring to nearby proteins [17]. The Chl "head" is formed of three 5-membered pyrrole rings and one partially saturated pyrrole (pyrroline) ring connected by methylene bridges. These four rings form a conjugated pi-orbital system which absorbs visible light. The presence of the pyrroline ring removes 4-fold symmetry from the chromophore and results in a shorter length of the conjugated pi-orbital system in the direction of the Q_x axis (Figure 2-2). This asymmetry causes different magnitudes of polarizations to occur along Q_y and Q_x axes. This is reflected in different magnitudes and energies of ground to singlet excited state transitions in the red part of the spectrum (so called Q_x and Q_y transitions), which are discussed in the next section.

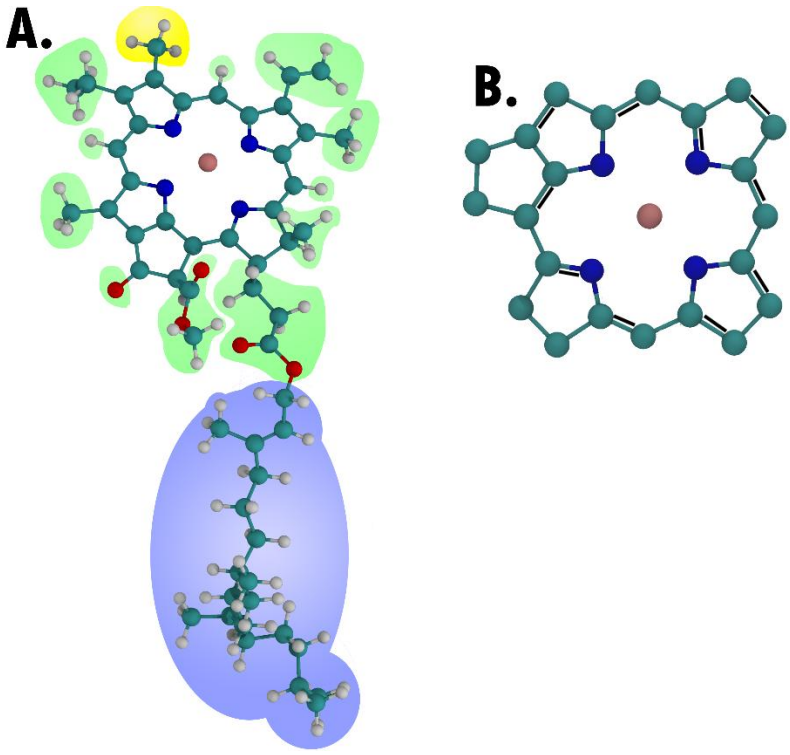

Figure 2-1 - (A) Chlorophyll a structure with phytyl chain or "tail" highlighted in blue, side chains highlighted in green and the side chain CH3 (side chain differentiating Chl a from Chl b) in yellow. (B) chlorin ring or "head" of a chlorophyll, with double bonds illustrated

2.2.2. Chlorophyll Excited States

Chl absorbs photons in the violet-blue and red regions of the visible spectrum as is shown in Figure 2-3.**Error! Reference source not found.** Absorption in the redmost band with a peak around 660 nm is due to transition from the ground state S0 to the first excited state S1 known as the Q_y transition [18]. Absorption in the second last band with a peak around 600 nm corresponds to $S0 \rightarrow S2$ transition which is known as a Q_x band, and absorption in the blue high energy band corresponds a number of higher, closely spaced transitions ($S0 \rightarrow S_n, n > 2$) known as the Soret-band [16]. Both Q_x and Q_y transitions induce polarization in the head group of the

Chl, each in a direction approximately orthogonal to one another in the plane of the head as shown in Figure 2-2. The first excited state causes a polarization in the Qy transition direction. Excited state energy levels of Chl embedded in a media (protein + other cofactors + membrane + solvent) depend on chromophore conformation, presence of external electric field, polarization of media by electron distribution of a ground state of a chromophore and polarization response of the embedding media to the chromophore transition to excited state [19]. Thusly the protein environment of a Chl can have a very strong effect on the Chl's absorption and hence energy transfer properties. The Qy excited state energy of a Chl in a protein environment is usually referred to as its "site energy". As excitation energy transfer in multi-chromophore photosynthetic systems depends on site energies, their knowledge is very important for understanding of this process.

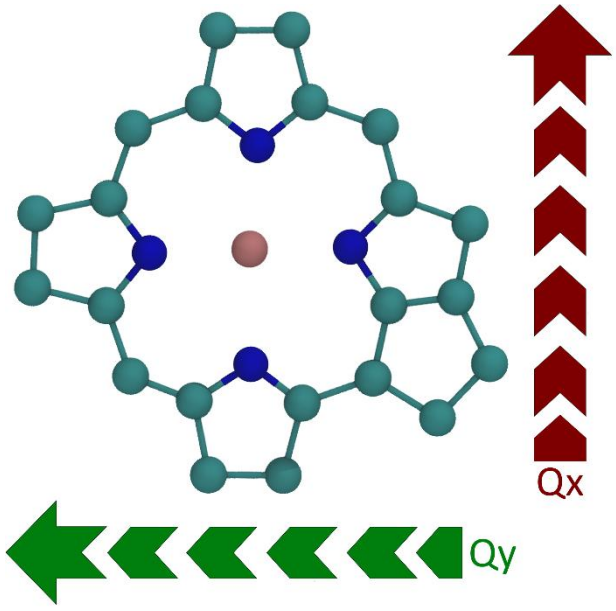

Figure 2-2 - Chlorophyll head group, with Qx and Qy transition directions.

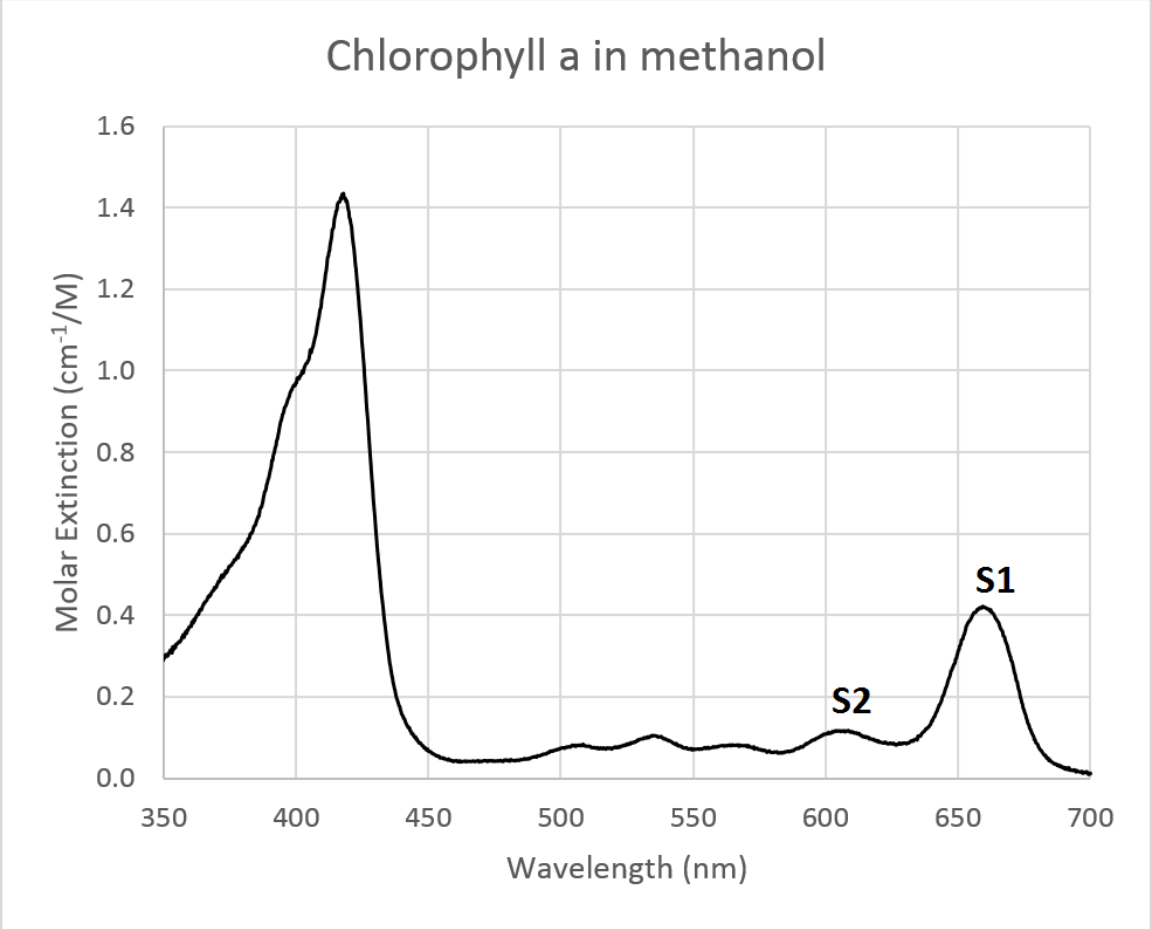

Figure 2-3 - Chlorophyll a's absorption spectrum (Dixon, Taniguchi, & Lindsey, 2005).

2.3. Excitation Energy Transfer In Photosynthesis

2.3.1. Energy Transfer between Fluorescent Chromophores

Excitation is transferred between fluorescing chromophores via non-radiative dipole-dipole coupling [20]. The rate at which energy is transferred is dependent on the relative positions, orientations and site energies of the donor and acceptor chlorophylls. Theory of resonance energy transfer was developed by Förster (FRET) [21]. According to FRET the rate of the energy transfer between two fluorescent chromophores is described by equation (2.1).

$$k_T = \frac{1}{\tau_D} \cdot \left(\frac{R_0}{r}\right)^6 \tag{2.1}$$

Where k_T is the rate constant of energy transfer from a donor molecule to an acceptor molecule. r is the distance between the donor and acceptor molecules and τ_D is the fluorescence lifetime of the donor in the absence of the acceptor. R_0 is the critical transfer distance (or Förster radius or Förster distance) (in units Å), which is the distance in which the transfer efficiency between the donor and acceptor is 50% and can be calculated using equation (2.1). k^2 is the orientation factor between the donor and acceptor given by equation (2.2), ϕ_0 is the quantum yield of the donor's fluorescence in absence of the acceptor, n is the refractive index of the intervening medium, J (units $M^{-1}cm^{-1}nm^4$) is the degree of spectral overlap between fluorescence spectrum and the acceptor absorption spectrum as seen in equation (2.3).

$$R_0 = 0.2108 \cdot [k^2 \cdot \phi_0 \cdot n^{-4} \cdot J]^{1/6} \tag{2.1}$$

The orientation factor k^2 given in equation (2.2) ranges in value from 0 (perpendicular dipoles) to 4 (collinear dipoles). $\hat{a}$ and $\hat{d}$ are the orientation factors for the acceptor and donor transition moments respectively. $\hat{r}$ is the displacement vector from donor to acceptor.

$$k^2 = \left(\hat{a} \cdot \hat{d} - 3(\hat{a} \cdot \hat{r})(\hat{r} \cdot \hat{d})\right)^2 \tag{2.2}$$

$$J = \int_0^\infty F_D(\lambda) \cdot \varepsilon_A(\lambda) \cdot \lambda^4 \cdot d\lambda \tag{2.3}$$

The overlap integral is given by equation (2.3), where λ is the wavelength, $F_D(\lambda)$ is the normalized emission spectrum of the donor and $\varepsilon_A(\lambda)$ is the molar attenuation coefficient. Lastly, Förster's transfer efficiency E given by equation (2.4) is dependent upon the distance (r) and transfer efficiency (R_0) of the donor and acceptor.

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \tag{2.4}$$

As overlap integral J depends on spectral overlap between donor and acceptor chromophores, site energy is one of the key factors controlling efficiency and the direction of propagation of excitation energy.

2.3.2. Site Energies of Photosynthetic Chromophores

It is crucial to understand the site energies of individual chromophores in order to be able to model excitation energy transfer in photosystems [22], [23]. Unfortunately site energy information is difficult, if not impossible to determine experimentally due to photosystems having multiple chromophores with overlapping spectra contributing to the experimentally observed spectra [22], [24]–[26]. Experimentally site energies can be determined by fitting the site energies to reproduce various spectral and kinetic data (absorption, fluorescence, linear dichroism, circular dichroism, transient absorption, fluorescence decay kinetics, e.t.c). However, in general, multiple sets of site energies may reproduce experimental data equally well. Due to this reason, computational approaches are required to validate and complement experimental results. Only structure-based calculations can relate differences in site energies to molecular features and shed light on how photosynthetic systems have evolved to achieve their unprecedented quantum efficiency.

2.4. Computation of Site Energies

As knowledge of the site energies of chromophores is crucial in understanding the intricacies and efficiencies of excitation energy transfer, calculations of site energies of chromophores using information from crystal structures has been attempted before. The excited state energy of a chromophore in vacuum can be computed using

several QM methods including semi-empirical method Zerner's Intermediate Neglect of Differential Overlap (ZINDO), time-dependent density functional theory (TDDFT), and configuration interaction singles (CIS). TDDFT is the most computationally efficient method that is known to give reasonable results for Chl. Although it overestimates the absolute value energies of excited states, it is capable of capturing differences between different conformations of chromophores and the effects of an external electric field.

Due to the computational complexity of QM, it is too costly to include all pigment-protein complexes in site energy calculations. To overcome the computational complexity of performing full quantum mechanical calculations on pigment-protein complexes, alternative methods for computing chromophore site energies have been devised. The earlier attempts have applied quantum mechanical techniques to only the chromophores and included some simplified model of their protein environment.

[27] Studied the effect of conformational distortions of the Chl macrocycle imposed by protein in the bacterial light-harvesting complex of P. aestuarii. This study provided insight into the effect of steric interaction with the protein on Chl site energies, however, it completely neglected effects of the protein electric field.

[28] Computed excited states of all 96 Chl in PSI in their X-ray conformations. To account for protein environment they included all residues within 2.5 Å of a Chl molecule in their QM calculations. With this approach any long-range electrostatic contributions to site energies from the rest of the protein were neglected. Another problem is that there are many cases where another chromophore is located within 2.5 Å of the Chl for which the site energy is being computed. Including such a

neighbor in the QM site energy calculation will yield excited states of a dimer, not a single Chl. Hence, such neighbors were neglected in this study.

[29] have done QM/MM (molecular mechanics) calculations of the excited states of individual chromophores in PSII, based on the molecular dynamics (MD) trajectories in order to determine the site energy of each chromophore, as well as how it changes with respect to time. To account for the chromophore environment, all atoms within 12 Å of any of the four Chl nitrogen atoms were included in the excited-state calculations as static point charges. This study concluded that X-ray coordinates of Chls are not sufficiently accurate for site energy calculations. Averaging of conformations from MD trajectories provided much better agreement with experimental spectra. However, this approach does not account for polarization of the protein/membrane/solvent environment.

The Charge Density Coupling method (CDC) was proposed to address this polarization problem. In CDC the difference between wave functions of excited and ground states is replaced by a set of atom-centered charges parameterized to reproduce the QM-derived change of electric field around Chl due to its excitation. This approximation allows for much faster calculation of Coulomb interactions between the electronic transition of the chromophore and the ground state charge distribution of the surrounding protein [4]. The CDC framework allowed the inclusion (implicit) polarization of protein dielectric volume and surrounding solvent or lipid membrane to the static electric field originating from neutral and charged amino acid side chains, the peptide backbone and other cofactors. As for any other approximation, CDC poses several shortcomings: it does not account for changes of site energies related to chromophore conformation, it does not account for polarization of the wave function by the ground state charges of the chromophore's local environment

and it does not account for dynamic polarization of the chromophore's environment caused by electronic transition from ground to excited state. These shortcomings could be problematic for site energy calculations. Thorough evaluation of the CDC method and point charge representation of protein environment indicated that both methods are not able to reproduce results of the reference Time-Dependent Density-Functional Theory (TD-DFT) calculations [5]. More rigorous treatment of the chromophore environment is necessary for reliable estimates of site energies.

Recently a method including both static and dynamic polarization of a protein environment into QM calculations of excited states has been developed [8]. Termed the polarizable embedding (PE) method, its purpose is to enable an environment to be able to induce effects into the electron density of a molecular core in a fashion in which the computational costs are low and accuracy is high. The PE method represents the environment of a molecular core via the permanent charge distribution of the environment by using a multicenter multipole expansion. These expansion centers are defined to be located at the atomic nuclei of the atoms and at the midpoints between atom bonds. To account for the polarization of the environment via itself and by the molecular core (represented quantum mechanically) a set of localized anisotropic dipole polarizability tensors are assigned to the expansion centers [30].

The core's environment, represented by multipole expansion centers which model its permanent charge distribution allows for the polarization of both the core and the environment. The response to the environments polarization from both itself and the core is included in the effective operator as induced dipoles (2.5). With the environments response computed, the electron density of the core is recomputed in a wavefunction optimization and the induced dipoles of the environment are once

again computed. It is in this method which PE is used to optimize the core's wave-function in a fully self-consistent manner. It is possible to obtain the multipoles and anisotropic dipole-dipole polarizability tensors in a multitude of ways, however, we will be using quantum mechanical methods. The PE model is implemented into density functional theory (DFT) by creating an effective Kohn-Sham (KS) operator $\hat{f}_{eff}$ which is expressed as the sum of the KS operator in vacuum $\hat{f}_{KS}$ and both the electrostatic and induced PE potential operators.

$$\hat{f}_{eff} = \hat{f}_{KS} + \hat{v}_{PE}^{es} + \hat{v}_{PE}^{ind} \tag{2.5}$$

The second quantized form of the expanded time-dependent KS Hamiltonian is given by

$$\begin{aligned} \hat{H}(t) &= \sum_{n} \hat{H}^{(n)} \\ &= \sum_{n} \sum_{pq} (\delta_{0n} h_{pq} + j_{pq}^{(n)} + \hat{v}_{xc,pq}^{(n)} + \hat{v}_{PE,pq}^{(n)}) \hat{E}_{pq} \end{aligned} \tag{2.6}$$

where h_{pq} is an integral over the kinetic energy and the nuclear attraction operators, $j_{pq}^{(n)}$ is an nth order Coulomb integral, $\hat{v}_{xc,pq}^{(n)}$ is an nth order integral over the exchange-correlation potential and $\hat{v}_{PE,pq}^{(n)}$ is the contribution from the polarizable environment given as an nth order integral over the PE potential [8].

2.5. Photosystem II Structure and Function

The core of PSII contains 18 identified transmembrane protein subunits, two of which (D1 and D2) form the reaction center located in the center of the photosystem [14]. The structure of PSII can be seen in Figure 2-4. Surrounding PSII's reaction center are antenna complexes containing antenna Chl whose purpose is to increase the

amount of photons which PSII captures by increasing the system's number of chromophores, the two proteins which form the structure for PSII's antenna complexes are known as CP43 and CP47. The reaction center of PSII contains 6 Chlorophylls and 2 Pheophytins. The photochemically active site of the reaction center resides at the pair of tightly coupled Chl molecules (D1/Chl604 and D2/ Chl605) called the special pair or P680. Two chlorophylls (D1/Chl 606 and D2/Chl 607) called "accessory chlorophylls" are found between P680 and the two pheophytins. The last two peripheral chlorophylls (D1/Chl 610 and D2/Chl 611) are involved in protection of the reaction center from oxidative damage in conditions when OEC cannot reduce P680. They are called chlorophylls Z.

Excitation energy from a photon trapped by a Chl in PSII will transfer its energy to a nearby Chl. The transfer is done through a non-radiative method and is repeated until the excitation energy reaches the PSII-RC [31]. From within the PSII-RC, the excitation is transferred to the special pair where a photochemical reaction occurs via electron transfer [32].

Once the special pair is excited to a higher energy level state, charge separation between P680 and D1 Chl 606 occurs. Charge separation is followed by electron transfer through the D1 subunit of the reaction center, from the special pair Chl 604 & 605 through Chl 606 to pheophytin 608 as seen in Figure 2-6, the pheophytin on the D1 subunit side. From pheophytin 608, the electron is transferred to a quinone electron acceptor in the D1 subunit known as Q_A. From Q_A the electron is then transferred to a secondary quinone acceptor in the D2 subunit known as Q_B. A negatively charged Q_B will then bind with a proton from the stroma. After this process occurs twice, the doubly reduced Q_B becomes a hydroquinone which has a

low binding affinity to the D2 protein and disperses off into the thylakoid membrane. A replacement quinone will bind itself in the freed location, becoming the new Q_B.

P680 has a very high oxidation potential, it has its electrons replenished each time it loses one via an electron coming ultimately from the oxygen evolving complex (OEC), which is comprised of four manganese, five oxygens and one calcium ion. In OEC water is oxidized within a manganese cluster (Mn_4O_5Ca cluster). With each electron transferred from the OEC to P680 the Mn_4O_5Ca cluster changes its oxidation state. There are four states which the Mn_4O_5Ca cluster cycles through known as the S states and they are labeled S_0 through S_4 [33]. The index in the S states represents the number of oxidizing equivalents stored in the Mn_4O_5Ca cluster [34]. With each electron transferred from the OEC to P680 the Mn_4O_5Ca cluster acquires an oxidizing equivalent changing its state from S_i to S_j, where j=(i+1) mod 5. State S_4 is an unstable state, and it is in this state where water is oxidized as the fourth oxidizing equivalent is accumulated. The electrons from the oxidized water replenish the OEC's lost electrons and the S_4 state quickly changes to the S_0 state. The protons from the water enter the lumen, contributing to a pH gradient across the thylakoid membrane and the oxygen molecules are released. As the total charge of the OEC is changing during its catalytic cycle the progression of S-states is expected to affect the excited states of RC chromophores.

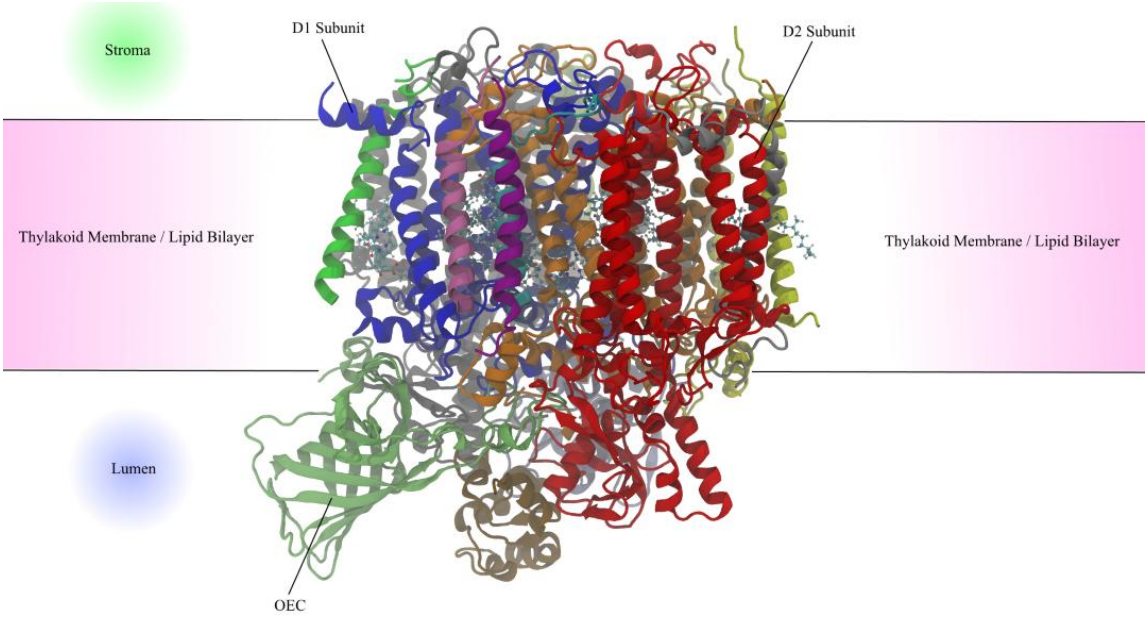

Figure 2-4 - PSII Protein Structure and 8 chromophore of the reaction center. The pink shaded area is the thylakoid membrane lipid bilayer. The area above the membrane is the stroma and the area beneath the membrane is the lumen. The reaction center proteins D1 and D2 are coloured blue and red respectively. The oxygen evolving complex is the protein structure in the bottom left corner.

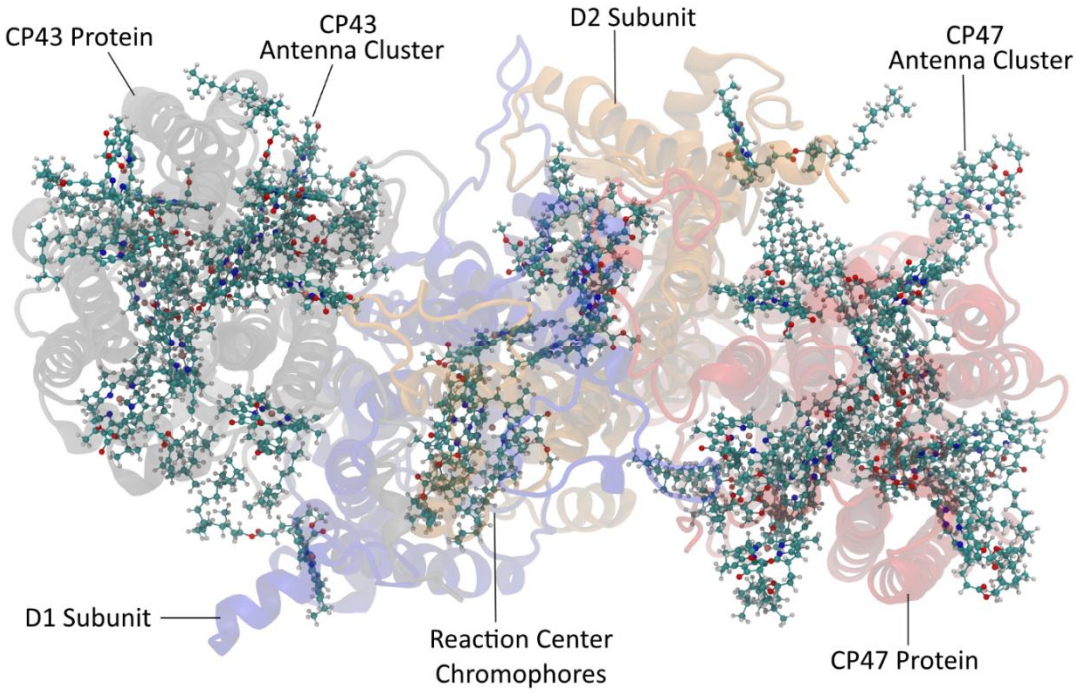

Figure 2-5 - PSII view from the stromal side of the lipid bilayer membrane. The protein structure is shown in feint as the background and illustrated molecules are the chlorophyll and pheophytin of the system. The clusters of molecules in the CP43 and CP47 protein complexes are the antenna chlorophylls. The chromophores clustered in the center in the D1 and D2 subunits for the chromophores of the reaction center.

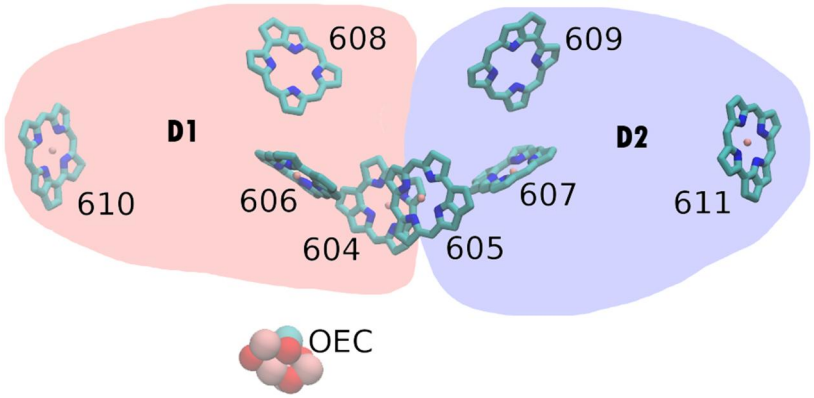

Figure 2-6 - Chromophores head groups of the PSII reaction center as oriented in the system. Chromophores in the light red region belong to the D1 subunit chain, chromophores in the light blue region belong to the D2 subunit. Chromophores 608 and 609 are pheophytin and all others are chlorophyll. Chromophores 604, 606, 608 and 610 belong to the D1 protein and chromophores 605, 607, 609 and 611 belong to the D2 protein. Chlorophyll 604 and 605 form the special pair where excitation energy is ultimately trapped. Chlorophyll 610 and 611 each sit on the outsides of the reaction center.

2.6. Photosystem II Reaction Center Site Energies

A remaining question with regards to PSII is the question as to what drives the electron transfer exclusively through the D1 subunit considering that D1 and D2 are near symmetric. Efforts have been made to try to elucidate the answer to this problem, however experimental efforts in determining the site energies of the PSII reaction center have produced disagreeable results. Experimentation has resulted in different relative site energies between the two pheophytins in PSII reaction centre. Using PSII core samples from spinach and Synechocystis measuring absorption and CD spectra at low-temperature, D2/Pho 609 (the pheophytin of the D2 subunit) has been found to have a higher energy level than D1/Pho 608 (the pheophytin of the D1 subunit) [35]. In these experiments two scenarios were run, in one Q_A had been

reduced to Q_A^- and the in other Q_A had been reduced to Q_A^- and then reduced again to form Q_AH_2. It was found that with Q_A^- absorption of D1 pheophytin is shifted 8 nm to the red relative to D2 pheophytin. This shift increased to 13 nm with Q_AH_2. In contrast, another study on the PSII reaction center preparations from spinach, wild-type Chlamydomonas reinhardtii and a Chlamydomonas reinhardtii mutant (D2-L209H) has results showing the D1 pheophytin to be blue shifted by 10 nm relative to D2 pheophytin [36]. The third study using PSII reaction centers from spinach reported that D1 and D2 pheophytins have approximately equal site energies [37]. The differences in site energies could be due to the different preparations (intact PSII core vs. isolated PSII-RC) used in these experiments. Only one study [35] used intact PSII core complexes, while the other two used PSII-RC (D1-D2-Cytb559 fragments of the PSII core). In this type of preparation the quinones which make up the pheophytin binding pocket in PSII core are lost. This may lead to a significant modification of pheophytin binding site and pheophytin conformation.

Though symmetric, the reaction center of PSII has been found to have some structural differences between the D1 and D2 chains [14]. Namely that the water ligand of D1-CLA606 is additionally bonded to a threonine (THR179) of the D1 protein through a hydrogen bond, a bond which does not exist in the analogous water ligand of D2-CLA607. Computational efforts have been used to investigate what effects the environment might have in creating differing site energies between the cofactors in the D1 chain and their analogous counterparts in the D2 chain. Using a QM/MM method where cofactors were treated quantum mechanically and their environment within 22 Å as point charges, it was found that D1-CLA606 site energy is downshifted by negatively charged OEC ligands [38].

3. Problem Definition

The goal of this project was to develop a computational framework for computing site energies of molecules in an environment in a fast and reliable fashion. Using this tool to address the following questions:

Why is the site energy of D1-CLA606 lower than that of its D2 counterpart?

Which of the two PSII-RC pheophytins has a higher site energy?

Do the state transitions of the OEC have a minimal effect on the site energy of D1-CLA606 due to the relative orientation of the two, even though they are close in proximity?

Furthermore, to use this tool to determine what protein features are defining the site energies and what dynamic effects the protein environment has on site energies and how it tunes the site energies in order to achieve the near unity quantum efficiency at which the reaction center operates at.

4. Methodology

This section describes the methodology involved in starting with an X-ray structure and producing site energies of desired chromophores. There are two problems which must be addressed.

The first problem concerns overcoming the relatively low resolution of existing X-ray structures. Information from current X-ray structures is often not detailed enough to provide sufficiently accurate information with regards to bond lengths, angles, and torsions between chromophore atoms in order to accurately compute site energies. To address the issue of X-ray accuracy, it is crucial to obtain an energy optimized ground state geometry of the chromophore in its protein environment. This is achieved via a computational approach in which small (up to 200 atoms) clusters are constructed where each cluster consists of the chromophore being optimized and its local (typically within 4 Å) environment. Each cluster is treated at QM level of theory. The environment is considered to be accurate enough as it is represented via MM, and is thus kept frozen during energy minimization. This process is described in further detail in section 4.1.

The second problem is the inclusion of the protein environment in calculations of site energies. We opted to use more rigorous representation of the environment than static point charges by including atom and bond centered polarizabilities, dipoles, and quadrupoles. This problem was tackled with two different approaches. The first solution was to fragment the environment around each chromophore into small QM systems, then compute localized wavefunction properties for each these fragments using the LoProp module of MolCas [39]. With the localized wavefunction peroperties, build a potential for each of the chromophoresto be used in their site energy

calculations. This is a computationally intensive solution which requires QM LoProp calculations to be performed for each residue/cofactor found around each chromophore. The eight chromophores of the RC have approximately 370 amino acids located within their 10 Å environment and each residue takes around 10-48 hours to complete on a quad-core 3.0 GHz Intel Core 2 CPU computer. These computations are too costly for calculating many site energies, thus an alternative approach less costly (computational time) approach is desired. A faster method for producing localised wavefunction properties of an environment has the added benefit of then also being able to calculate inhomogeneous broadening. With the current method, calculating inhomogeneous broadening on a 1 ps time scale with 1 fs time intervals would thus require two and a half to three years of computational time to perform on a single computer

An alternative approach was implemented to drastically reduce the computational time. With this second approach, we exploited the fact that the LoProp parameters of small rigid functional groups within molecules are independent of the overall molecular conformation. Therefore localized wavefunction properties of each type of residue/cofactor can be computed once and stored in a molecular mechanics database (usually called a force field). Then, to build a complete embedding potential each functional group from the molecular database is mapped (translated and rotated) to the corresponding group in the chromophores' environment. This approach requires only one QM computation for each type of molecule in the system to be done at the preparation stage. It completely eliminated repeated QM computations of the environment in studies of the effects of mutations and/or protein dynamics on site energies. The calculation of site energies is described in section 4.2.

4.1. Preparing Chromophores for Site Energy Calculations

4.1.1. Adding Hydrogens to Crystallographic Structure

A vast majority of x-ray structures do not include hydrogen molecules due to the high resolution required before hydrogen atoms can begin to be resolved. The highest resolution x-ray structure of PSII at the present time is 1.9 Å[14], which does not meet the resolution of 1 Å roughly required before hydrogen atoms can begin to be resolved. Thus the first step in the process of computing site energies from a starting x-ray structure is to add hydrogen atoms to the structure. This is performed using the Amber software package which holds definitions of residues that include all atom names which belong to a residue which it uses to determine the missing (hydrogen) atoms and places them by using the coordinates of the atoms in the structure. Amber uses molecular mechanics to relax the newly added hydrogens into low energy configurations.

4.1.2. Preparing Chromophore Environment for Optimization

In order to obtain high accuracy optimizations of the ground state structure of chromophores, it is important to include environmental effects. However in order to obtain results in a feasible amount of time, it is not possible to include the entire PSII structure when performing the ground state calculations for chromophores. For this reason only the environment in close proximity to the chromophore is included for ground state calculations. To ensure computations can be performed within an acceptable amount of time, keeping the local environment to a minimal size is crucial. Molecules were chosen to be part of a chromophore's local environment based on whether or not the distance between the molecule and chromophore fell within a certain cut off distance and the distance between the two molecules is defined as the distance between the two closest atoms from each molecule.

Additionally two measures were taken to insure the local environments be kept to a minimum size. The first measure is to avoid selecting entire proteins and only select individual amino acids falling within the cut off distance. To ensure chemical sensibility is maintained, selected amino acids are capped with ACE or NME as described in section 4.1.2.2. The second measure was to divide large cofactor molecules into smaller fragments and only select the fragments of molecules falling within the cut off distance, similar to how the individual amino acids were selected as opposed to the entire protein. Selected fragments are capped with hydrogen atoms using the same method as in section 4.1.1 in order to preserve chemical sensibility. An example of a Chl and its local environment with a 4.0 Å cut off distance can be seen in Figure 4-1.

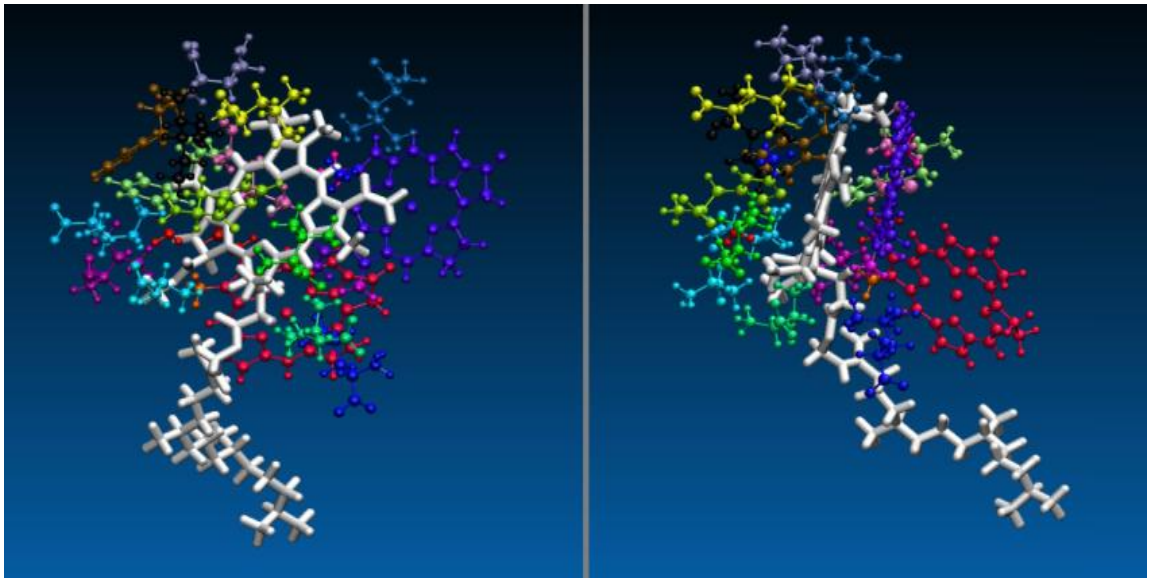

Figure 4-1 - A partition of PSII with a Chl (white molecule) head as the core of the partition and a cut off distance of 4Å used for selecting the local environment.

4.1.2.1. Fragmenting Large Cofactors

Minimizing a chromophore's local environment is desirable in order to reduce computational times of ground state optimizations. Many times when selecting molecules for a local environment there will be large molecules which are included, yet only have a very small portion of their structure within the cut off distance. The inclusion of such a residue can greatly increase the number of atoms in a local environment where most of those atoms will have a negligible effect on the chromophore's optimization due to their distance. Contrarily not including such a residue is also not desirable as accuracy of the optimization would diminish due to missing atoms that are in close proximity. Our solution to this problem is to "break" large molecules into fragments and include all molecule fragments which have an atom that falls within the cut off distance.

An example of how molecules were broken into fragments can be seen in Figure 4-2, which demonstrates how beta-carotenes were fragmented and the partitioning of all cofactors can be seen in the appendix in section 8.5. The cofactors were fragmented at single carbon bonds, where each fragment would contain roughly the same number of atoms. Simply selecting fragments for an environment is not sufficient as chemical sensibility would not be preserved. Using the beta-carotene in Figure 4-2 as an example, if the blue fragment were to fall within the cut off distance of a core chromophore, while the orange fragment did not the blue fragment would exist in the cluster in a nonsensical way. In order to maintain chemical sensibility a hydrogen is added to the selected fragment in a fashion that replaces the missing bond between the selected fragment and the missing fragment.

The Amber software package was used to manually define residue fragments. The definitions included the fragments unique name, atoms names as they appear in the original X-ray structure PDB file (section 8.2) and would also include the capping hydrogens. Before selecting local environments, the PDB file containing the X-ray structure of PSII was processed, renaming residue fragments from their original residue names to their fragment names.

After selection of a chromophore's initial environment, the selected molecules and molecule fragments are loaded into the Amber software package which adds the terminating hydrogens to all of the residue fragments. Since terminating hydrogens are added to all residue fragments regardless of whether or not two connected fragments are present, all extra hydrogens added to fix the broken bonds between two residue fragments that have both been selected are removed. In order to determine if multiple residue fragments from the same molecule have been selected and hydrogens need to be removed, all residues in the environment are checked and for each residue fragment found the file is then checked for all residue fragment types which come from the same molecule, and for each of these found, the distance between the two capping hydrogens is checked to see if their distances are close in order to ensure that the two fragments were from the same molecule and not just the same molecule type.

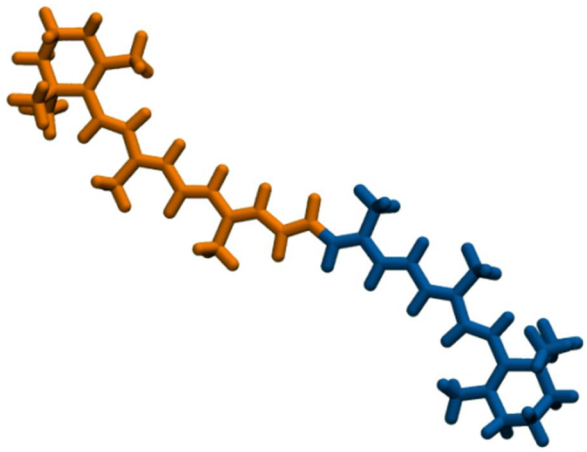

Figure 4-2 - Beta-carotene, illustrating the two partitions used for environment selection. The beta-carotene was split into two partitions shown by the orange and blue colouring. If any atom from the orange partition fell within selection range then the entire orange part would be included in that environment, likewise with the blue partition. If at least one atom from each partition falls within a selection range then the entire beta-carotene is selected.

4.1.2.2. Protein Selecting

Proteins can be very large molecules, the D1 protein in PSII consists of 5360 atoms and the D2 protein contains 5144 atoms. The inclusion of an entire protein molecule in a local environment is thus not a viable option.

Proteins are generally too large to include in a quantum mechanical calculation desired in a feasible amount of time on modern computers, and so must be handled with extra care in order to prevent partition sizes from being too vast. To solve this issue, proteins are handled in a similar fashion to the large molecules that were

partitioned as described in section 4.1.2.1. For proteins however, instead of specifying how and where the molecule will be broken into fragments, each amino acid composing the molecule is treated as a fragment, or its own molecule.

Amino acids are selected to be in a partition if one of their atoms falls within the cut-off distance of an atom within one of the core molecules of that partition in the same fashion regular molecules are selected for partitions. When an amino acid is selected for a partition and one or both of its neighboring amino acids are not selected, then like a partially selected fragmented residue, the amino acid must be capped in order to maintain chemical sensibility. When an amino acid is capped, it is capped by either an acetyl beginning group (ACE) if terminating the C-terminus side or an N-methylamine ending group (NME) if terminating the N-terminus side. NME and ACE caps are used instead of the regular terminating sequences seen in Figure 8-1 as the NME and ACE more closely resemble the peptide bonds between the amino acids.

Additionally there are cases where a single amino acid is not present in a partition and its two neighboring amino acids are present. In these scenarios, the NME and ACE added to the two neighbours of the removed amino acid will not have enough space to coexist in a natural fashion. To resolve this issue, instead of terminating the two amino acids with ACE and NME, the missing amino acid is brought into the partition, having its R group replaced by a hydrogen, turning it into a glycine. This preserves chemical sensibility, while keeping the number of atoms in the partition low.

NME and ACE caps are added to an amino acid (Acid) by using the position of three atoms from either the amino acid preceding Acid (bAcid) or the amino acid proceeding Acid (nAcid) or three atoms from each. When terminating with ACE, three atoms from bAcid are used, carbon “C”, oxygen “O” and alpha carbon (CA), where CA is renamed to “CH3”. When terminating with an NME, nAcid is used for its three atoms, nitrogen “N”, hydrogen “H” and alpha carbon “CA’ which is renamed to “CH3”. These six atoms from neighbouring amino acids can be seen in Figure 4-3, which illustrates an amino acid capped by both NME and ACE.

To add the remaining hydrogens to both NME and ACE, the structure is loaded into Amber’s LEaP program (see section 8.3), which contains definitions for the full structures of both NME and ACE. Using the coordinates of the six atoms provided by bAcid and nAcids a reference adds the hydrogen atoms missing from both NME and ACE. The added hydrogen can be seen in Figure 4-3, they are the unmarked atoms in the NME and ACE boxes.

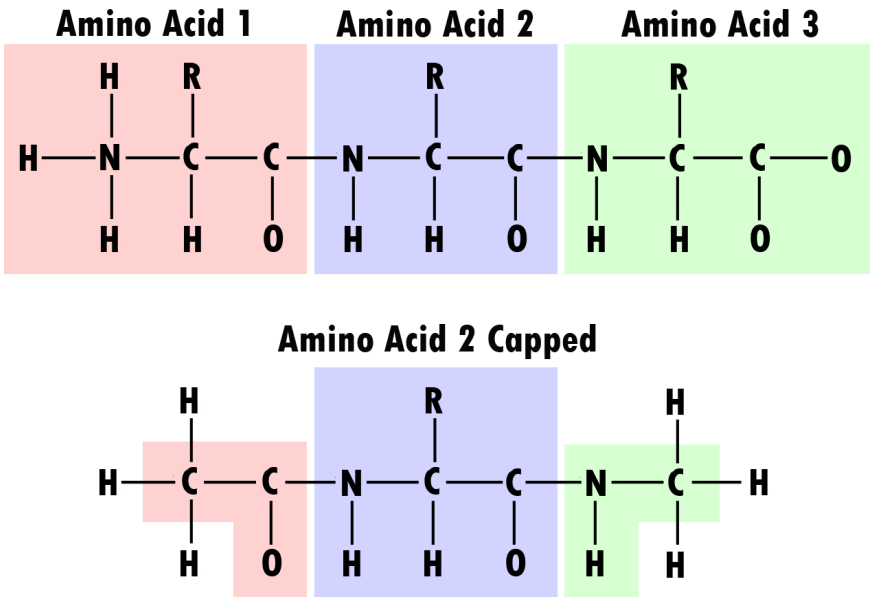

Figure 4-3 - Capping an amino acid. (Top) Three amino acids chained together. (Bottom) Amino Acid 2 from top separated and capped. The two carbons and oxygen highlighted in light red on the left are atoms taken from Amino Acid 1 who are used by LEaP to add and position the three hydrogens on the left who are not highlighted. The nitrogen, carbon and hydrogen highlighted in light green on the right are three atoms taken from Amino Acid 3 and are used by LEaP to add and position the three hydrogen atoms on the right who are not highlighted.

4.1.3. Optimizing Chromophore Geometry

As site energies are defined by the difference between energy levels of the ground state and the first excited state, it is crucial to use highly accurate equilibrated ground states of molecules when computing excited state properties. To achieve this high level of accuracy in a reasonable amount of time a hybrid of two different quantum mechanical techniques were implemented using Gaussian software package's ONIOM method. With this method, the chromophore being optimized is treated at a higher level using Density Functional Theory (DFT) while the chromophore's environment was handled with a semi-empirical quantum mechanical method. The X-ray structure of the environment was also assumed to be accurate enough and the environment was fixed in place for the optimizations. The entire PSII structure is

too large to run with such semi-empirical method and thus only part of the chromophore's environment was selected (as described in section 4.1.2) for inclusion in relaxation calculations. To aid in keeping environment sizes small, large molecules were "broken" into fragments so that large molecules with only a small number of atoms falling within the cut off distance would only have their fragments which are within the cut off distance included. When only a fragment of a residue falls within the cut off distance it is capped by hydrogen atoms to ensure chemical sensibility. The process of fragmenting residues is described in section 4.1.2.2. Lastly due to the immense size of proteins, it is crucial to not include an entire protein with an atom within the cut off distance of the core. To overcome this, amino acids were considered individually for selecting the environment opposed to the entire protein molecule. Selected amino acids whose bonding amino acids were not selected were capped using ACE and/or NME to ensure chemical sensibility was maintained. The process of selecting amino acids is described in greater detail in section 4.1.2.2.

4.2. Calculating Site Energies with Multipoles & Polarizabilities

When computing site energies with the inclusion of protein environment and cofactors, we opted to use a more rigorous approach than simply representing the environment as static point charges and instead represented the environment as multipoles and polarizabilities (as visualised in Figure 4-4) located at the environment's expansion centers. To compute localized wavefunction properties, the local environment of each chromophore was selected in the same fashion as described in section 4.1.2. Multipoles were computed up to and including the 2^{nd} order, consisting of charge, dipoles and quadrupoles as well as the polarizabilities. These wavefunction properties were computed using the LoProp module of MolCas and two different

approaches were devised for computing and applying the properties to the local environment.

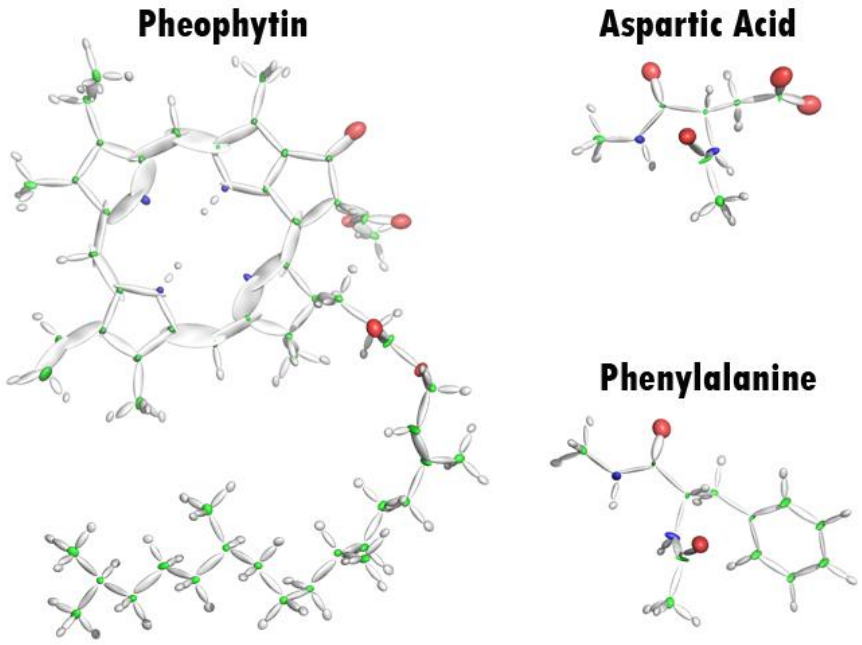

Figure 4-4 – Visualisation of polarizabilities of Pheophytin, Aspartic Acid and Phenylalanine. Polarizabilities are located at atom centers and atom bond midpoints. Hydrogen and bond centers are white, nitrogen are blue, carbon are green and oxygen are red.

4.2.1. Method I – Merging All Files

The first method used was to compute the localized wavefunction properties of each residue in the local environment. Cofactors and amino acids were separated into individual files (amino acids were capped with NME and ACE as described in section 4.1.2), and the properties were computed using the LoProp module of MolCas and the results were merged together to form the environment in which the chromophore site energies would be computed. The computed wavefunction properties were then brought together into a single file. Bringing cofactors is as simple as concatenating a list of properties, amino acids require more work as atoms in ACE and NME had

their properties computed twice (once as terminal atoms of an amino acid in ACE or NME and a second time during the neighboring amino acid's computations). Additionally, amino acids also have extra hydrogen caps as part of their NME and ACE which are not present in the structure of the local environment, which must be removed. Thusly the ACE and NME of two amino acids bonded in the local environment are removed. The total charge of the ACE is added to the last carbon of the preceding amino acid and the total charge of the NME is added to the first nitrogen of the proceeding amino acid (as seen in Figure 4-5).

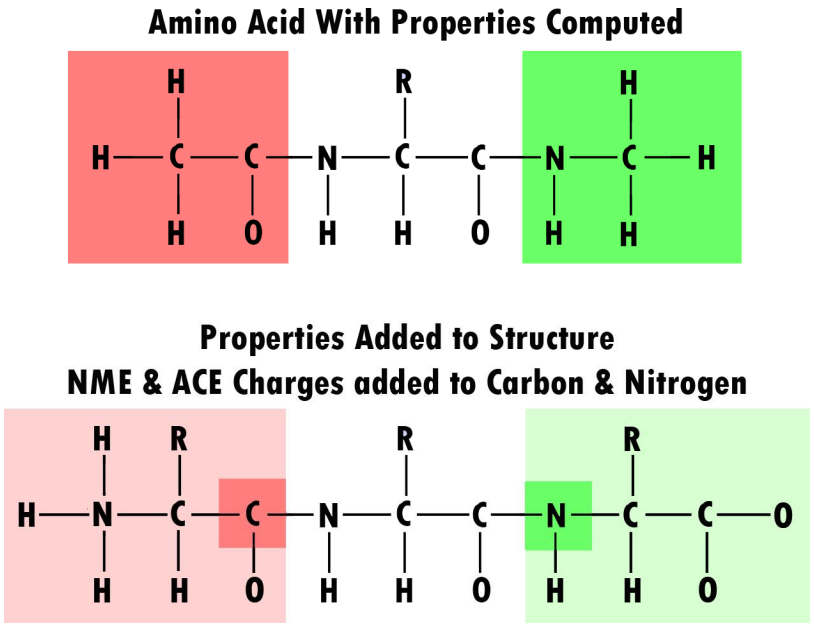

Figure 4-5 – The amino acid (top) has its computed charges and wavefunction properties applied to the amino acid in structure (bottom). The charges of non ACE and NME atoms are applied directly to the atoms of the amino acid in structure. The total charge of the ACE (dark red top) is added to the charge of the last carbon (dark red bottom) of the preceding amino acid (light red bottom) in the chain. The total charge of the NME (dark green top) is added to the charge of the first nitrogen (dark green bottom) of the proceeding amino acid (light green bottom) in the chain.

4.2.2. Method II – Fitting Potentials to a Structure

The goal of method II was to achieve the wavefunction properties of a chromophore's local environment with the same accuracy as method I, but much more quickly. Unlike method I, which requires computations of each individual residue in the local environment to have its wavefunction properties computed, method II only requires that each type of residue have its wavefunction properties computed a single time. The results from the single residue type are then applied to all residues of that type within the local environment. This method takes advantage of the conformation of a residue having negligible impact on the values of its wavefunction properties other than their orientations. With the wavefunction properties computed using LoProp module of MolCas of each residue type, the properties are then fitted to all of the residues of the same type in the local environment. The advantage of this method increases when storing the results of the wavefunction properties for each residue type in a library, meaning that they do not need to be computed when building a local environment. Since fitting the properties to a structure takes such little time it results in the ability of going from selecting an environment to computing site energies in relatively no time.

To fit the wavefunction properties to the conformations of the residues in the local environment, the residues in the local environment had the atom bond midpoints computed and added to the structure. The expansion centers were then formed into groups in a way so that each group would contain at least three atoms and their two bond midpoints (more expansion centers if the three atoms form a straight line) and that each expansion center belongs to at least one group. Corresponding groups were formed from the wavefunction properties and each of these groups were fitted to their corresponding structural group.

Fitting the groups of wavefunction properties to the corresponding groups of the structure is performed using quaternions to calculate the root-mean-square deviation (RMSD)[40]. This is achieved via finding an orthogonal transformation U for each group, which minimizes the equation:

$$E = \frac{1}{N}\sum_{i=1}^{N} |U\overline{p_i} - \overline{s_i}|^2 \tag{4.1}$$

Where N is the number of atoms in the group, p is the set of coordinates of the wavefunction properties to be rotated and s is the set of coordinates of the atoms in the group of the structure being fit to. To minimize E we represent our orthogonal transformation U in quaternion form. This requires writing $\overline{p_i}$ and $\overline{s_i}$ as quaternions $\dot{p}_i$ and $\dot{s}_i$ where $\dot{p}_i = (0, \overline{p_i})$ and $\dot{s}_i = (0, \overline{s_i})$. Writing equation (4.1) in quaternion form yields:

$$E_q = \frac{1}{N}\sum_{i=1}^{N} (\dot{q}\dot{p}_\iota\dot{q}^* - \dot{s}_i)(\dot{q}\dot{p}_\iota\dot{q}^* - \dot{s}_i)^* \tag{4.2}$$

Where $\dot{q}$ is our rotation quaternion in which we now wish to solve for in order to minimize our residual E_q. Expanding and rearranging (4.2) yields:

$$E_q = \frac{1}{N}\left[\sum_{i=1}^{N} (|\overline{p_i}|^2 - |\overline{s_i}|^2) - 2\dot{q}\mathcal{F}\dot{q}^T\right] \tag{4.3}$$

where $\mathcal{F}$ is given by:

$$\mathcal{F} = -\sum_{i=1}^{N} \mathcal{R}_l\,(\dot{s}_i)\,\mathcal{R}_r(\dot{p}_\iota) \tag{4.4}$$

And $\mathcal{R}_l(\dot{q})$ and $\mathcal{R}_r(\dot{q})$are 4x4 matrices given by (4.5) and (4.6).

$$\mathcal{R}_r(\dot{q}) = \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & q_3 & -q_2 \\ q_2 & -q_3 & q_0 & q_1 \\ q_3 & q_2 & -q_1 & q_0 \end{pmatrix} \tag{4.5}$$

$$\mathcal{R}_l(\dot{q}) = \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix} \tag{4.6}$$

$\mathcal{F}$ can be expanded to

$$\mathcal{F} = \begin{pmatrix} P_{11} + P_{22} + P_{33} & P_{23} - P_{32} & P_{31} - P_{13} & P_{12} - P_{21} \\ P_{23} - P_{32} & P_{11} - P_{22} - P_{33} & P_{12} + P_{21} & P_{13} + P_{31} \\ P_{31} - P_{13} & P_{12} + P_{21} & -P_{11} + P_{22} - P_{33} & P_{23} + P_{32} \\ P_{12} - P_{21} & P_{13} + P_{31} & P_{23} + P_{32} & -P_{11} - P_{22} + P_{33} \end{pmatrix} \tag{4.7}$$

where $P_{ij} = \overline{p_i} \cdot \overline{s_j}$. This reduces the problem of minimising the residual to solving the eigenvalue problem

$$\mathcal{F}\dot{q}^T = \lambda\dot{q}^T \tag{4.8}$$

The largest eigenvalue will pose the solution to minimizing the residual. The Jacobi method was used to diagonalise matrix $\mathcal{F}$ and determine the rotation quaternion. The left rotation matrix U is then obtained from the quaternion by applying the quaternion rotation to the 3x3 identity matrix.

$$U(\dot{q}) = \dot{q}I\dot{q}^T \tag{4.9}$$

$$U(\dot{q}) = \begin{pmatrix} \dot{q}_0{}^2 + \dot{q}_1{}^2 - \dot{q}_2{}^2 - \dot{q}_3{}^2 & 2(\dot{q}_1\dot{q}_2 - \dot{q}_0\dot{q}_3) & 2(\dot{q}_1\dot{q}_3 + \dot{q}_0\dot{q}_2) \\ 2(\dot{q}_1\dot{q}_2 + \dot{q}_0\dot{q}_3) & \dot{q}_0{}^2 - \dot{q}_1{}^2 + \dot{q}_2{}^2 - \dot{q}_3{}^2 & 2(\dot{q}_2\dot{q}_3 - \dot{q}_0\dot{q}_1) \\ 2(\dot{q}_1\dot{q}_3 - \dot{q}_0\dot{q}_2) & 2(\dot{q}_2\dot{q}_3 + \dot{q}_0\dot{q}_1) & \dot{q}_0{}^2 - \dot{q}_1{}^2 - \dot{q}_2{}^2 + \dot{q}_3{}^2 \end{pmatrix} \tag{4.10}$$

The rotation matrix $U(\dot{q})$is then applied to the wavefunction properties of a group so that they fit the structure. Finally, the negative of the translation vector used to center the structure's centroid on the origin is applied to the wavefunction properties to translate them to their final positions which they will take in the structure representation. Just as in method I, there is an issue of atoms in the amino acid backbone having their wavefunction properties computed multiple times as some atoms belong to two amino acids. This was handled in the exact same fashion as was described in section 4.2.1.

4.2.3. Computing Site Energies

A 1.9 Å resolution crystal structure of PSII [14] was prepared as described in section 4.1. Polarizabilities and multipoles of the structure were computed using the methods described in section 4.2. Site energies of chromophores were computed using the Dalton molecular electronic structure program's implementation of the polarizable embedding method described in section 2.6. A visualisation of a Dalton input file can be seen in Figure 4-6, with PSII represented as point charges and multipole expansions for everything within 10 Å of the chromophore being inspected. The work flow for computing site energies of a chromophore can be seen in Figure 4-7.

One of the input files required by Dalton is a list of atom types and coordinates and bond-midpoints and coordinates, preceded by lists of charges, multipoles and polarizabilities. These properties are computed as described in either section 4.2.1 or 4.2.2. Lastly the input file also contains an exclusion list. The exclusion list has an entry for each atom and bond-midpoint in the file and each entry contains a list of all atoms and bond-midpoints which cannot polarize the current entry. For cofactors, the exclusion lists of the atoms and bond-centers in a given molecule would have exclusion lists which contain all other atoms and bond-centers in the same molecule.

This is due to the way in which polarizabilities are computed. Polarizabilities are computed using MolCas via applying a linear electric field to the molecule six times, in both directions of three orthogonal vectors. The polarizabilities of each site in the molecule are computed while taking into account the effects of all other sites (atoms and bond-centers) as they are affected by the same fields. As such, the polarizabilities of a residue are computed together and sites of the same residue should not apply their effects from polarizations to the other sites of that residue, as this would magnify the effect of the field inducing the polarization, which is why Dalton uses the exclusion lists.

The exclusion lists for sites in amino acids in proteins require special treatment, as they are covalently bonded forming polymer chains. As the covalent bonds linking amino acids are replaced by conjugated caps during polarizability computations, exclusion lists for the amino acids sites must be extended to include sites from neighbouring amino acids. Thus exclusion lists of sites in amino acids include all sites of neighbouring amino acids within a distance of three covalent bonds in addition to all other sites of the same amino acid. Example of exclusion lists of amino acid sites can be seen in Figure 4-8.

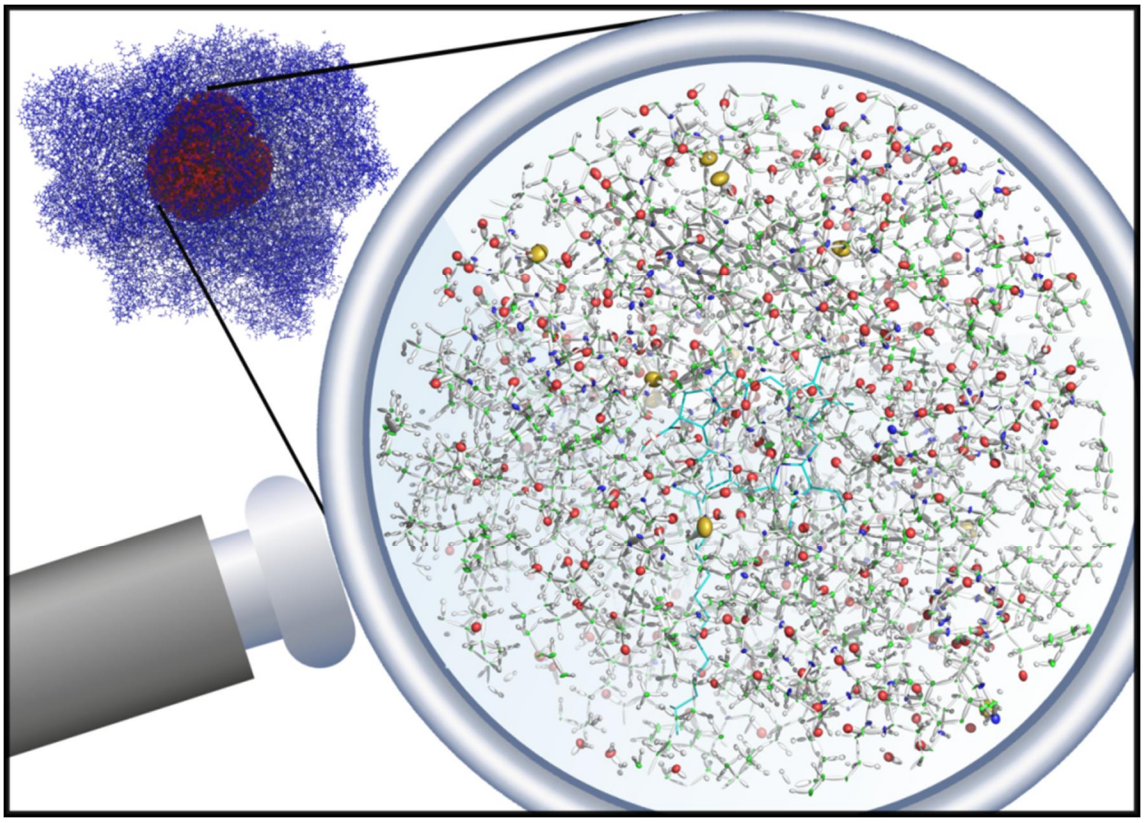

Figure 4-6 - Top left, entire PSII structure in blue as atoms and bonds with an example 10A local environment of a sample Chl in red. Bottom right shows a magnification of the red volume from the top left and shows polarizabilities from local wavefunction property calculations. The core Chl can be seen in light blue in the magnification with the Chl head located in the center of the magnification.

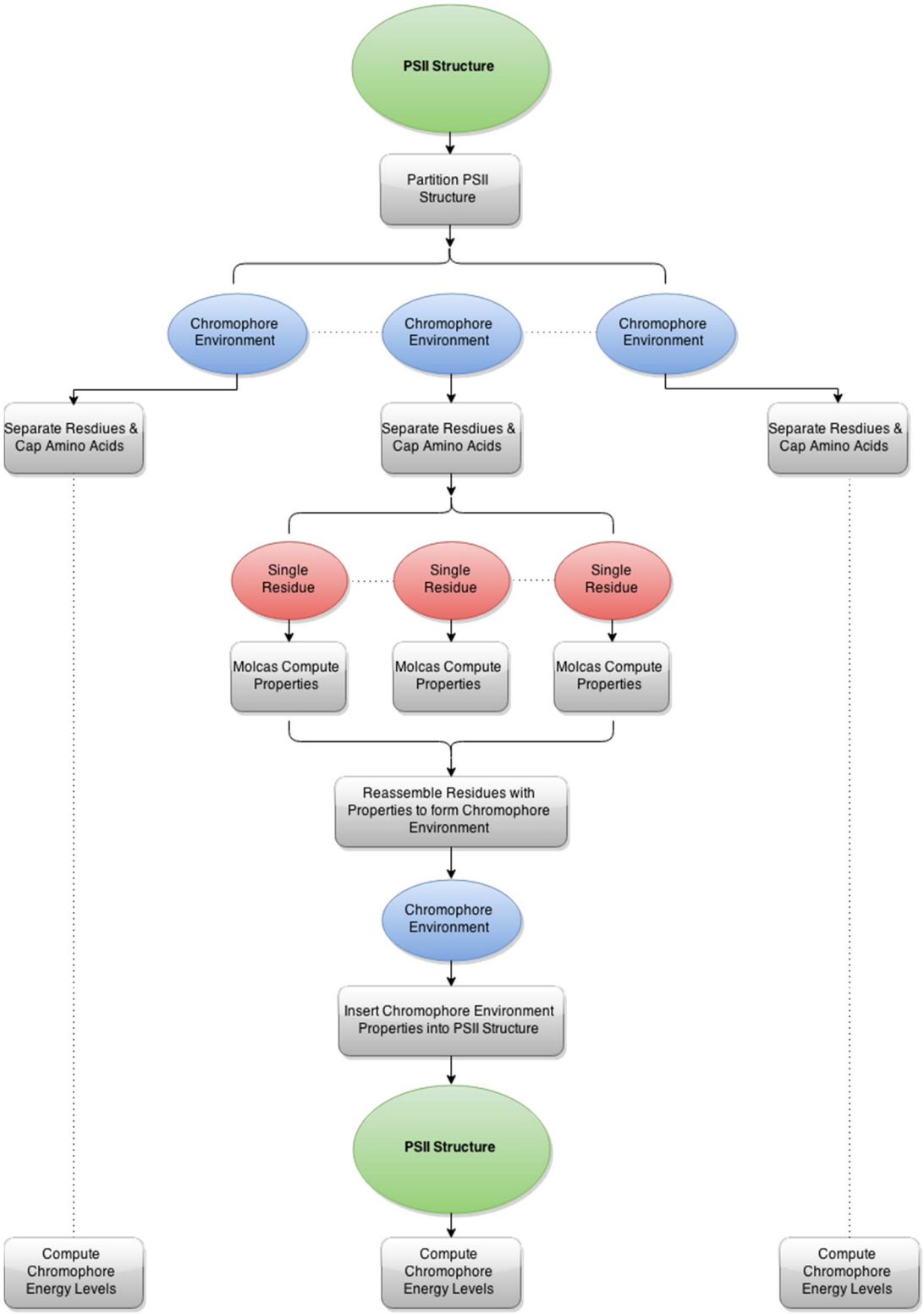

Figure 4-7 -Work flow for computing site energies of chromophores

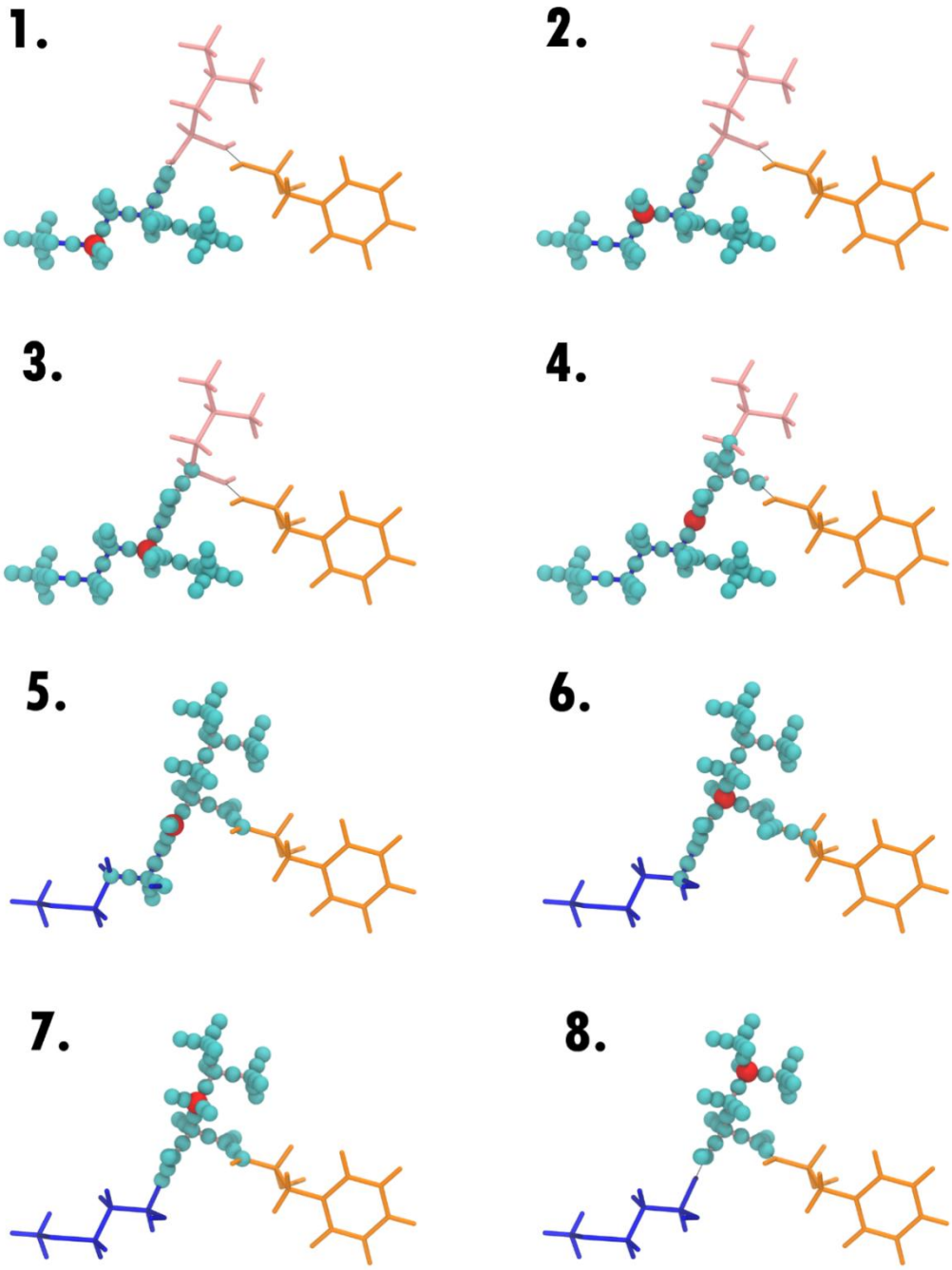

Figure 4-8 - Amino acid exclusion lists. Each image (1-8) shows the sites (turquoise spheres) of the selected atom's (red sphere) exclusion list as well as the structure of three neighboring amino acids (MET (blue), LEU (pink) and PHE (orange)). Image 1. The selected atom of MET is too far from the LEU to have any sites excluded from LEU. Images 2-4 show atoms in MET closer to LEU have more atoms from LEU in their exclusion lists. Images 5-8 show exclusion lists for atoms in LEU and so contain all sites in LEU while containing sites from MET and PHE depending on the atoms bond distance from them.

.

5. Results and Discussion

A 1.9 Å resolution crystal structure of PSII [14] was prepared by performing QM energy optimization (as described in section 4.1) on the eight RC chromophores using a 4.0 Å cut off distance for their local environments. The optimizations were performed using Gaussian's ONIOM, with the chromophores being treated with functional b3lyp and basis set 6-31g* as they are commonly thought to produce good results for optimizing Chl a. Their environments were handled with PM3. The site energies of the eight RC chromophores were computed using their QM energy optimized conformations while including different amounts of information from their environments. As a means of comparison, the results from these computations were compared to PSII-RC chromophore site energies obtained via a simultaneous fit of absorption, linear dichroism, circular dichroism, fluorescence and Stark spectra of PSII-RC obtained at low temperature (4-10K) and 77 K [41].

5.1. Site Energies of Reaction Center Chromophores

Site energies of the RC chromophores were computed using varying levels of chromophore optimization and representation of their environment during the computations. These levels included:

(1) the unoptimized chromophores taken directly from the x-ray crystallographic structure, site energies computed in vacuum (X-ray);

(2) the ground state optimized chromophores, site energies computed in vacuum (Vacuum);

(3) the optimized chromophores, site energies computed with a point charge representation of chromophore environment for all atoms within 10 Å of the chromophore (10A-M0);

(4) the optimized chromophores, site energies computed with a point charge representation of the entire PSII structure (All-M0);

(5) the optimized chromophores, site energies computed with first and second order multipoles within 10 Å of the chromophore in addition to a point charge representation of the entire PSII structure (All-M0&10A-M1-M2);

(6) the optimized chromophores, site energies computed with a point charge representation of the entire PSII structure and first and second order multipoles as well as first order polarizabilities within 10 Å of the chromophore (All-M0&10A-M1-M2-P1).

The results from these computations can be seen in Table 5-1. As DFT generally overestimates site energies, the computed values were higher than the experimental. The inclusion of polarizabilities caused a significant drop in the site energies, bringing them closer to the experimental results than that of the other computations with less information in the environments.

The site energy of D1/Chl 610 was reported to be higher than that of D2/Chl 611 from the experimental fits [41]. However, as these two Chls are situated near opposite edges of the PSII-RC, distant from other chromophores within the RC, swapping site energies of these Chls would not affect quality of simulated spectra. Therefore experimental site energies of D1/Chl 610 and D2/Chl 611 are near indistinguishable from one another. This opens the possibility that in contrast to previous reports,

D1/Chl 611 has a higher site energy than D1/Chl 610 as was found in All-M0&10A-M1-M2 and All-M0&10A-M1-M2-P1 calculations.

There is a discrepancy with the two Pheos site energies between computed results and experimentally fitted results [41]. Experimentally the two Pheos were found to have lower site energies relative to the Chls, with D1/Chl 606 being an exception as it was found to have the lowest site energy of all, a feature in agreeance with previous work using hole-burning spectroscopy [42]. This is in contrast to our computational results which found D1/Pho 608 to have the highest site energy of the RC chromophores. D2/Pho 609 was also found to have a high site energy relative to the other Chl. The origin of this discrepancy is not clear at the moment. We cannot see any molecular features which could lower site energies of Pheos more than that of Chls, and this is reflected in our computations. D1/Pho 608 having a higher site energy than D2/Pho 609 is also in agreeance with previous work [37].

Table 5-1 - Site energies (1/cm) of the eight chromophores of the PSII-RC determined through fitting of experimental data [41] and different computational techniques. The highest, second highest, lowest and second lowest site energies of each run are marked in the following colours:

highest	second highest	second lowest	lowest

Chromophore Number:	604	605	606	607	PHO608	PHO609	610	611
Experimental Fitting	15190	15180	15000	15130	15050	15060	15555	15485
X-Ray	16101	16419	16259	16208	15981	15965	15927	15788
Vacuum	17248	17220	17331	17404	17436	17314	17432	17448
10A-M0	17183	17250	17253	17379	17591	17423	17442	17428
All-M0	17271	17277	17171	17370	17529	17410	17454	17426
All-M0 & 10A-M1-M2	17333	17277	17113	17384	17522	17407	17441	17511
All-M0 & 10A-M1-M2-P1	16436	16447	16518	16834	17143	16998	16877	17034

One feature of the relative site energies observed in the experimental fitted results is that D1/Chl 606 has the lowest site energy of the eight chromophores, making it the

energy trap of the PSII-RC. This is a feature also found in our computations in both All-M0 and All-M0&10A-M1-M2. However when polarizabilities are included, both D1/Chl 604 and D2/Chl 605 were found to have lower site energies than D1/Chl 606. The lowering of their site energies could be due to inadequate treatment of these two strongly coupled Chl during their energy optimizations and site energy calculations. For future work, further investigation is required to determine if treating D1/Chl 604 and D2/Chl 605 as a dimer in energy optimization and/or site energy computation would produce a better match of the experimental results.

5.2. Comparison of the computational methods.

As a means of assessing quality of the different computational methods, we will now look at the differences between experimental and computed energy levels. We will use two criteria: (1) root-mean-square deviation between computed and experimental site energies, and (2) the spread of site energies.

For the reasons described above we have excluded the P680 dimer and both Pheos from calculations of RMSD. The results are shown in Figure 5-1. As seen in this Figure site energies computed using Chl conformations taken directly from crystallographic data have very poor predictive power (rmsd=430 1/cm). QM cluster energy optimization significantly improved computed site energies even without inclusion of any environment model (rmsd=265 1/cm). Addition of static electric field represented by multipole expansion of orders 0, and 0-2 progressively improved quality of calculated site energies (rmsd=125 1/cm). The best result was achieved when along with multipoles up to 2^{nd} order dipole-dipole polarisation was included (rmsd=115 1/cm).

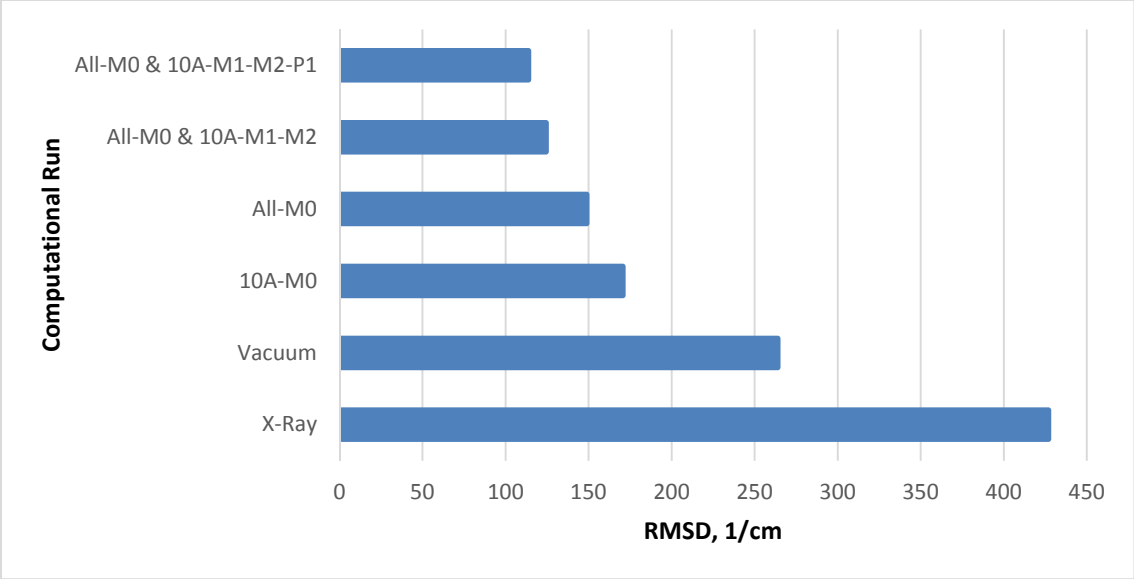

Figure 5-1 - RMSD between computed and experimentally fitted site energies [41]. See text for details.

Next we computed differences in the spread of site energy levels. We define the spread to be the difference between the maximum of D1/Chl 610 and D2/Chl 611 (the two highest site energies from experimental fitting) and D1/Chl 606 (the lowest site energy from experimental fitting). Figure 5-1 shows the difference in spread between the experimentally determined results and the computational results. Here we can see again that the inclusion of polarizabilities (All-M0&10A-M1-M2-P1) produces a spread which most closely matches the spread of the experimental site energies.

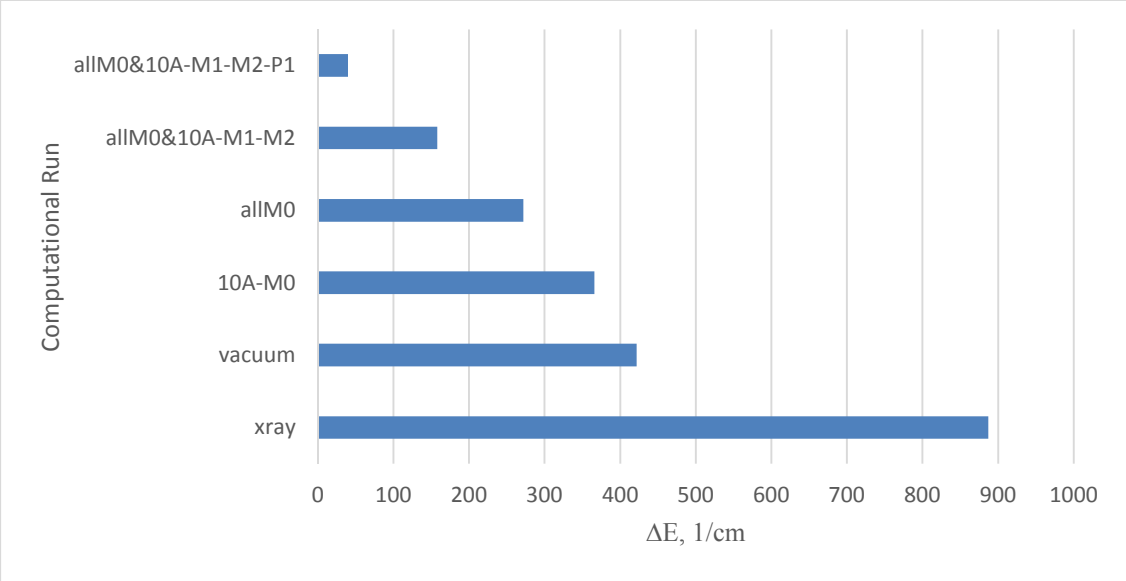

Figure 5-2 - The difference in spread between the experimentally fitted site energies [41] and the site energies computed using different environment model, see text for details.

5.3. Effect of OEC on D1-CLA606

Site energies of the eight RC chromophores were computed with the OEC configured in each of its five S - states. A 1.9 Å resolution crystal structure of PSII [14] with chromophores optimized as described in section 4.1 and a point charge representation of the PSII environment was used. The results shown in Figure 5-3 indicate that the state of the OEC complex has negligible effects on the energy level of D1-CLA606. This is in direct contradiction with a previous computational study which reported that the D1-CLA606 site energy is downshifted by negatively charged OEC ligands [38]. These results indicating the OEC has a great influence on D1-CLA606 were surprising and not reasonable due to the orientation of the head group of D1-CLA606 relative to the OEC. As can be seen in Figure 2-6, the S0-S1 transition of the D1-CLA606 is almost perpendicular to the vector running from the center of the head

group to the center of the OEC and thus the electric field from the OEC would have a very small component parallel to the Q_y direction.

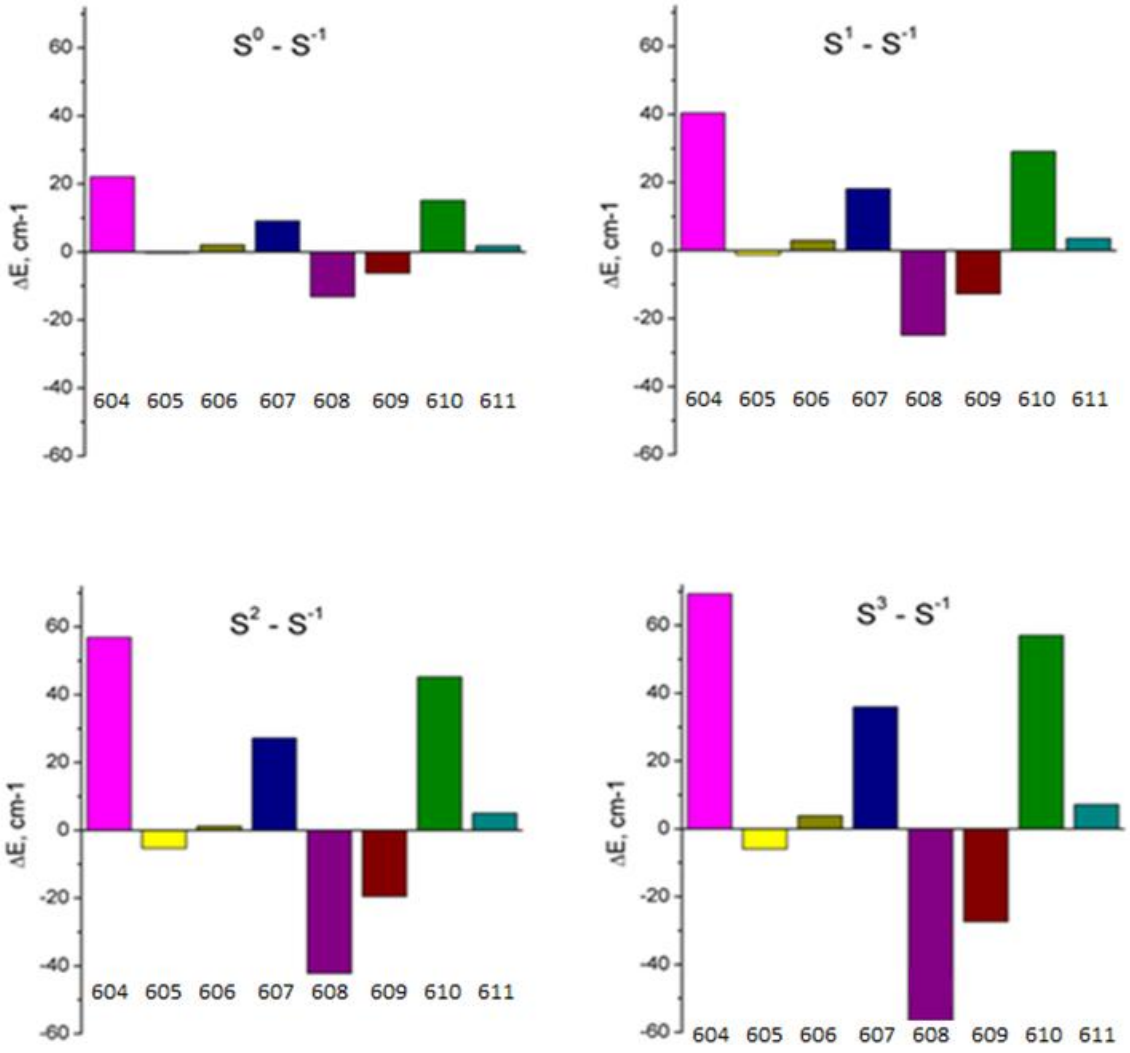

Figure 5-3 - Energy level differences of Reaction Center chromophores between OEC's S^{-1} and its other states (S^0 – S^3).

5.4. D1-PHO608 and D2-PHO609 Relative Energy Levels

Experimental results of the site energies of D1/Pho 608 and D2/Pho 609 are not in agreement with one another. Using PSII-RCs purified from spinach it has been found

that D1/Pho 608 and D2/Pho 609 have identical site energies [37]. Later, it has been found using full PSII core complexes from spinach that the site energy of D1-PHO608 is about 175 1/cm higher than the site energy of D2/Pho 609 [35]. The latest study of isolated PSII-RC from wild-type Chlamydomonas reinhardtii placed D1/Pho608 220 1/cm lower than its D2 counterpart D2/Pho 609 [37].

Our results show that D1-PHO608 has a site energy which exceeds that of D2/Pho 609 by 168 1/cm (CIS) and 206 1/cm (TD). This result is in agreement with the experimentalists who used intact PSII core preps. We found that the difference in the pheophytins site energies can be partially attributed to conformational differences between them. These conformational changes are due to differences in their environment, differences which stem from the D1 and D2 proteins. The pheophytins ground state conformations were optimized using the methods described in section 4.1.3, using B3LYP/6-32G** for the pheophytins and PM3 for their environments which consisted of all residues within 4 Å of their macrocycles. The differences in conformation between D1/Pho 608 and D2/Pho 609 can be seen in Figure 5-4.

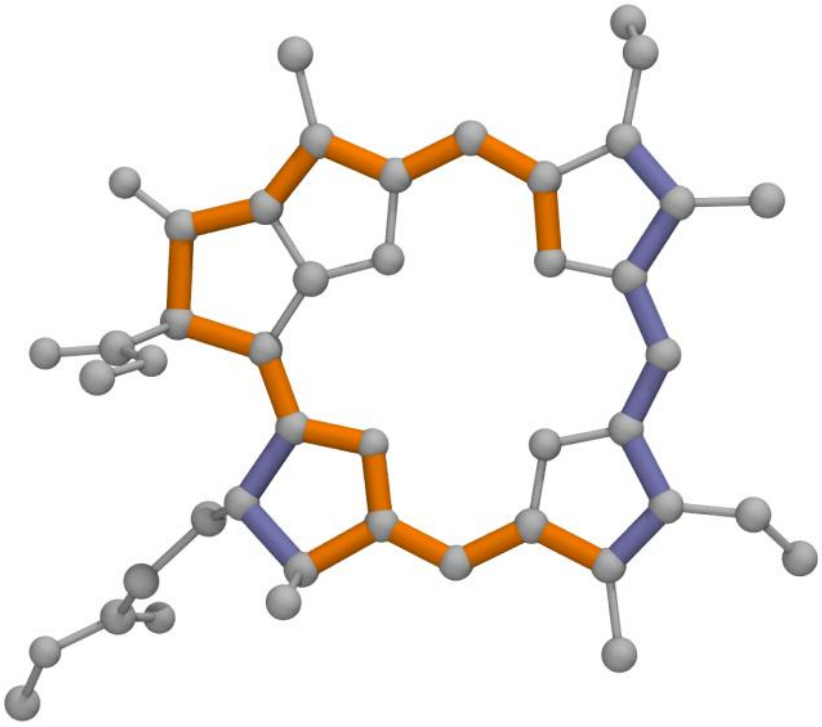

Figure 5-4 - Differences in optimized ground state conformations of D1-PHO608 and D2-PHO609. The blue lines indicate bond lengths which are longer in D1-PHO608 and the orange lines represented bond lengths which are longer in D2-PHO609.

To determine the direct contribution of the environment to the differences between D1/Pho 608 and D2/Pho 609, the site energies of each of the two pheophytins were computed in vacuum, in the presence of one of each of their five surrounding amino acids. The energy shifts due to the environmental factors can be seen by comparing the site energies to the site energies computed in vacuum. The site energies were computed using both Configuration Interaction Singles (CIS) and Time-Dependent Density-Functional Theory (TD-DFT) and the direction of the energy shifts due to environmental factors were in agreement. The results of these calculations can be seen in Table 5-2 where it can be noted that the energy shifts with all five amino

acids included is not equal to the sum of the energy shifts of the five calculations where a single amino acid was included.

Table 5-2 - Energy level (1/cm) differences of D1/Pho 608 and D2/Pho 609 between results computed in vacuum and with single amino acid in their environment and five amino acids in their environment. Computations were performed with CIS and TD.

PHO608	**NONE**	**ILE213**	**ILE143**	**GLN130**	**PHE252**	**TYR126**	**ALL 5**
CIS	18533	-66	-61	-39	-59	-70	-59
TD	17071	-80	-73	-52	-71	-84	-76
PHO609	**NONE**	**MET214**	**ASN142**	**GLN129**	**TYR254**	**PHE125**	**ALL 5**
CIS	18412	-102	-58	-48	-54	-73	-105
TD	16951	-118	-73	-63	-67	-89	-130

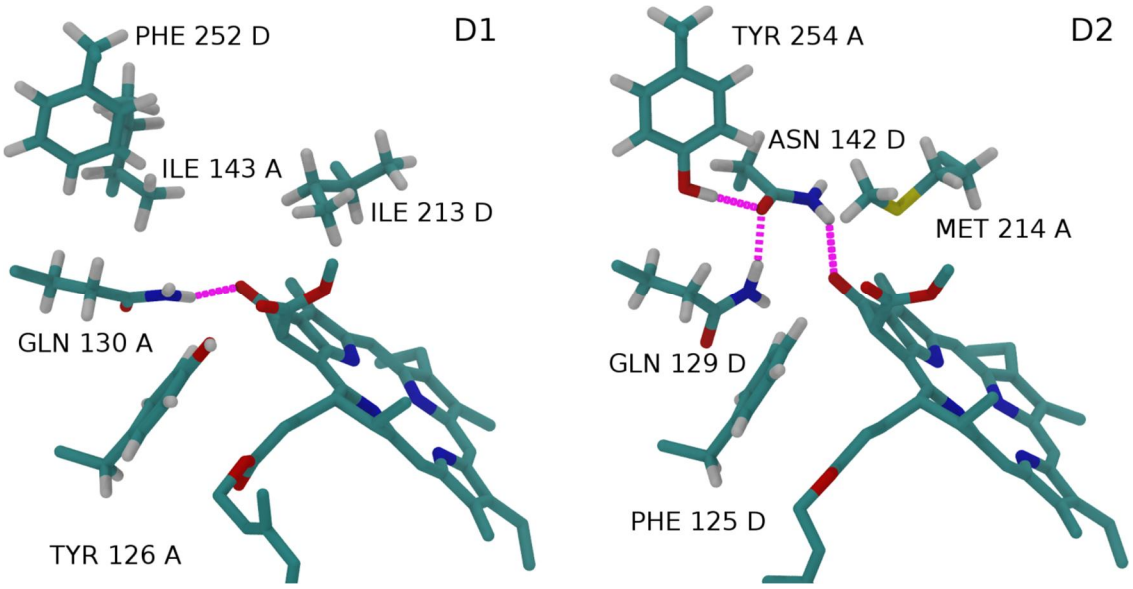

Figure 5-5 - Protein around D1-PHO608 and D2-PHO609. Five amino acids were included in QM calculation of excited states.

To summarize: the conformational changes were found to be responsible for about 75% of the energy difference between D1/D1 Pheo. Another 25% of site energy shift is due to direct effect of Pheo environment. These results indicate that chromophore environments have a large potential to affect site energies indirectly.

6. Conclusion and Future Work

A framework was developed for the automatic generation of ONIOM input to perform cluster QM energy optimization of chromophores. The framework produces an ONIOM input file for each desired chromophore. Each of the generated inputs include the structure of the chromophore being optimized with the chromophore's local environment and the input dictates a high level of QM theory to use on the chromophore and a lower level of QM theory for the environment. The framework involves breaking down larger molecules into smaller components to allow for smaller environments for faster QM energy optimizations with minimal sacrifice to accuracy. This framework was successful in allowing for ONIOM input to be created with a sufficiently sized local environment which would run in a sufficient time frame of about 24 hours on a quad-core 3.0 GHz Intel Core 2 CPU computer.

A second framework was developed to quickly produce input files for site energy calculations. These input files included the polarisable environment potential which would be constructed in one of two ways. For the residues in the local environment, if the polarisable potential already exists for the same or different conformation of the residue then the polarizabilities are fitted to match the conformation of the residue as it exists in the input file. This is a fast process taking advantage of quaternions (section 8.4) and an input file consisting entirely of residues with existing polarizabilities can be constructed within fractions of a second. This is opposed to the 10-48 hours it takes to compute the polarizable potential of a single residue on a single quad-core 3.0 GHz Intel Core 2 CPU computer. To ensure the fast creation of input files for computing site energies polarisable potentials were developed for all cofactors found in PSII and all amino acids, including their N- and C- terminal

occurrences. With this framework it is possible to investigate the role protein dynamics have in controlling site energies, as the framework enables the computation of site energy calculations for large ensembles of conformations arising from MD simulations in feasible amounts of time.

Performing site energy computations on the chromophores when including all point charges of the PSII structure, first and second order multipoles of all atoms within 10 Å of the chromophores and all polarizabilities of atoms and bond centers within 10 Å of the chromophores lead to successful convergence of the site energies via the Dalton software package in a reasonable time of less than one day. Extending the polarizabilities to include all atoms and bond centers within 15 Å of the chromophores resulted in systems too large to be handled on modern computers. The 15 Å environments drastically increased the computation time to unfeasible times and led to convergence failures.

Application of the developed framework to calculation of site energies in PSII lead to several interesting results. We have shown that QM optimization of chromophore in its protein environment is essential to obtain meaningful site energies. ONIOM optimization eliminated the unreasonable spread of computed site energies of the RC chromophores arising from the low precision of crystallographic conformations. In addition it showed that chromophore conformation is a significant factor influencing site energies that cannot be neglected.

Our calculations matched the difference in site energies between D1/D2 pheophytins determined using complete PSII core preparations [35], and we were able to assign this difference to a set of several amino acids in PSII reaction center. This result

strongly suggests that pheophytin site energies in isolated D1/D2/cytb559 preparations are modified by biochemical purification procedures. One of the likely explanations is direct modification of pheophytin binding sites by plastoquinone extraction during the isolation procedure. Our ONIOM calculations indicated that the protein environment controls site energies of pheophytins by modifying both electronic structure and nuclear geometry, with both of these mechanisms acting synergistically. This complicates understanding of the mechanism of site energy shift, and further work is required to identify major factors influencing site energies. To this end a detailed quantification of contributions from static atomic multipoles and oxidation polarizabilities is required.

Calculations of site energies of the RC with different oxidation states of the OEC revealed that site energy of the lowest energy reaction center chlorophyll (accessory Chl D1) is not affected by accumulation of positive charge in the process of water oxidation. This result clearly shows that previous assignment of the accessory Chl D1 site energy shift to OEC ligands [38] is incorrect. Thus, molecular features responsible for lowering of its site energy remain elusive. This question is well suited for the framework developed in this thesis and will be addressed in future studies. In addition the dependence of site energies on the oxidation state of the OEC provides a basis for modeling excitation transfer and trapping efficiency in PSII at all steps of its physiological reaction cycle.

Our initial calculations of site energies using only static atomic multipoles underestimate the difference between the reaction center core chlorophylls and the peripheral D1/D2 chlorophylls Z. Inclusion of atomic polarizabilities brings the computed difference close to experimental and even somewhat overestimates it. This result is expected as each of the reaction center chlorophylls has at least two other highly

polarizable chromophores in its close environment, while the peripheral chlorophylls do not. Interestingly, our PE calculations place D1 ChlZ lower than its D2 counterpart. We assigned this difference to a highly polarisable carotenoid which is in van der Waals contact with D1 ChlZ. Fitting of the experimental results also required different site energies for D1/D2 ChlZ, however in contrast with our calculations D1 ChlZ was placed higher than its D2 counterpart [41]. We note that as chlorophylls Z are weakly coupled to the rest of RC chromophores swapping their site energies would not affect the quality of the fit.

Unexpectedly while experimental studies [41] place pheophytins at the lower end of the spectra, our calculations did not reproduce this, and we were unable to find any molecular features that could be shifting pheophytins site energies down. It is possible that higher level excited state calculations are required for more accurate comparison of Pheo with Chl.

With this framework, which provides the ability to compute site energies of chromophores both quickly and accurately it is now feasible to perform time dependent simulations of either electron transfer or excitation energy transfer through any photosystem while incorporating quantum mechanical theory. These simulations would involve computation of protein and cofactor dynamics through MM computations accompanied by QM computations of chromophore site energies every n^{th} frame of MM intervals.

7. Bibliography

[1] A. Tanaka and A. Makino, "Photosynthetic research in plant science," *Plant Cell Physiol.*, vol. 50, no. 4, pp. 681–683, 2009.

[2] B. Loll, J. Kern, W. Saenger, A. Zouni, and J. Biesiadka, "Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II," *Nature*, vol. 438, no. 7070, pp. 1040–1044, Dec. 2005.

[3] W. Saenger, P. Jordan, and N. Krauß, "The assembly of protein subunits and cofactors in photosystem I," *Curr. Opin. Struct. Biol.*, vol. 12, no. 2, pp. 244–254, Apr. 2002.

[4] F. Müh, M. E.-A. Madjet, J. Adolphs, A. Abdurahman, B. Rabenstein, H. Ishikita, E.-W. Knapp, and T. Renger, "Alpha-helices direct excitation energy flow in the Fenna Matthews Olson protein.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 43, pp. 16862–16867, 2007.

[5] S. Vassiliev, A. Mahboob, and D. Bruce, "Calculation of chromophore excited state energy shifts in response to molecular dynamics of pigment-protein complexes.," *Photosynth. Res.*, vol. 110, no. 1, pp. 25–38, Oct. 2011.

[6] R. E. Blankenship, D. M. Tiede, J. Barber, G. W. Brudvig, G. Fleming, M. Ghirardi, M. R. Gunner, W. Junge, D. M. Kramer, A. Melis, T. a Moore, C. C. Moser, D. G. Nocera, A. J. Nozik, D. R. Ort, W. W. Parson, R. C. Prince, and R. T. Sayre, "Comparing photosynthetic and photovoltaic efficiencies and recognizing the potential for improvement.," *Science*, vol. 332, no. 6031, pp. 805–809, 2011.

[7] X.-G. Zhu, S. P. Long, and D. R. Ort, "Improving photosynthetic efficiency for greater yield.," *Annu. Rev. Plant Biol.*, vol. 61, pp. 235–261, 2010.

[8] J. M. Olsen, K. Aidas, and J. Kongsted, "Excited States in Solution through Polarizable Embedding," *J. Chem. Theory Comput.*, vol. 6, no. 12, pp. 3721–3734, Dec. 2010.

[9] J. Losos, K. Mason, and S. Singer, *Biology*, 8th ed. New York: McGraw-Hill, 2008.

[10] M. K. Sener, S. Park, D. Lu, A. Damjanovic, T. Ritz, P. Fromme, and K. Schulten, "Excitation migration in trimeric cyanobacterial photosystem I," *J. Chem. Phys.*, vol. 120, no. 23, pp. 11183–11195, Jun. 2004.

[11] S. Vasil'ev, P. Orth, A. Zouni, T. G. Owens, and D. Bruce, "Excited-state dynamics in photosystem II: Insights from the x-ray crystal structure," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 15, pp. 8602–8607, Jul. 2001.

[12] R. Nevo, D. Charuvi, O. Tsabari, and Z. Reich, "Composition, architecture and dynamics of the photosynthetic apparatus in higher plants," *Plant J. Cell Mol. Biol.*, vol. 70, no. 1, pp. 157–176, Apr. 2012.

[13] R. E. Blankenship, "Early evolution of photosynthesis.," *Plant Physiol.*, vol. 154, no. 2, pp. 434–8, Oct. 2010.

[14] Y. Umena, K. Kawakami, J.-R. Shen, and N. Kamiya, "Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å," *Nature*, vol. 473, no. 7345, pp. 55–60, May 2011.

[15] P. Jordan, P. Fromme, H. T. Witt, O. Klukas, W. Saenger, and N. Krauss, "Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution.," *Nature*, vol. 411, no. 6840, pp. 909–917, 2001.

[16] H. Scheer, *Chlorophylls*. CRC Press, 1991.

[17] P. H. Raven, R. F. Evert, and S. E. Eichhorn, "Photosynthesis, Light, and Life," in *Biology Of Plants*, W.H. Freeman, 2005, pp. 119–127.

[18] A. Marin, I. H. M. Van Stokkum, V. I. Novoderezhkin, and R. Van Grondelle, "Excitation-induced polarization decay in the plant light-harvesting complex LHCII," *J. Photochem. Photobiol. Chem.*, vol. 234, pp. 91–99, 2012.

[19] K. Sneskov, T. Schwabe, J. Kongsted, and O. Christiansen, "The polarizable embedding coupled cluster method," *J. Chem. Phys.*, vol. 134, no. 10, 2011.

[20] D. Gust, T. A. Moore, and A. L. Moore, "Molecular mimicry of photosynthetic energy and electron transfer," *Acc. Chem. Res.*, vol. 26, no. 4, pp. 198–205, Apr. 1993.

[21] T. Förster, "Zwischenmolekulare Energiewanderung und Fluoreszenz," *Ann. Phys.*, vol. 437, no. 1–2, pp. 55–75, Jan. 1948.

[22] J. Adolphs, F. Müh, M. E.-A. Madjet, M. S. am Busch, and T. Renger, "Structure-Based Calculations of Optical Spectra of Photosystem I Suggest an Asymmetric Light-Harvesting Process," *J. Am. Chem. Soc.*, vol. 132, no. 10, pp. 3331–3343, Mar. 2010.

[23] G. Raszewski and T. Renger, "Light harvesting in photosystem II core complexes is limited by the transfer to the trap: can the core complex turn into a photoprotective mode?," *J. Am. Chem. Soc.*, vol. 130, no. 13, pp. 4431–4446, 2008.

[24] K. Hingorani, B. Conlan, W. Hillier, and T. Wydrzynski, "Elucidating Photochemical Pathways of Tyrosine Oxidation in an Engineered Bacterioferritin 'Reaction Centre,'" *Aust. J. Chem.*, vol. 62, no. 10, pp. 1351–1354, 2009.

[25] T. Renger and E. Schlodder, “Primary Photophysical Processes in Photosystem II: Bridging the Gap between Crystal Structure and Optical Spectra,” *ChemPhysChem*, vol. 11, no. 6, 2010.

[26] M. Wendling, M. Przyjalgowski, D. Gülen, S. Vulto, T. Aartsma, R. van Grondelle, and H. van Amerongen, “The quantitative relationship between structure and polarized spectroscopy in the FMO complex of Prosthecochloris aestuarii: refining experiments and simulations,” *Photosynth. Res.*, vol. 71, no. 1, pp. 99–123, 2002.

[27] E. Gudowska-Nowak, M. D. Newton, and J. Fajer, “Conformational and environmental effects on bacteriochlorophyll optical spectra: correlations of calculated spectra with structural results,” *J. Phys. Chem.*, vol. 94, no. 15, pp. 5795–5801, 1990.

[28] A. Damjanović, H. M. Vaswani, P. Fromme, and G. R. Fleming, “Chlorophyll Excitations in Photosystem I of Synechococcus elongatus,” *J. Phys. Chem. B*, vol. 106, no. 39, pp. 10251–10262, Oct. 2002.

[29] S. Vasil’ev and D. Bruce, “A Protein Dynamics Study of Photosystem II: The Effects of Protein Conformation on Reaction Center Function,” *Biophys. J.*, vol. 90, no. 9, pp. 3062–3073, May 2006.

[30] J. M. H. Olsen and J. Kongsted, “Chapter 3 - Molecular Properties through Polarizable Embedding,” in *Advances in Quantum Chemistry*, vol. 61, J. R. S. and E. Brändas, Ed. Academic Press, 2011, pp. 107–143.

[31] R. van Grondelle, “Excitation energy transfer, trapping and annihilation in photosynthetic systems,” *Biochim. Biophys. Acta*, vol. 811, pp. 147–195, 1985.

[32] J. Berg, J. Tymoczko, and L. Stryer, “Accessory Pigments Funnel Energy Into Reaction Centers,” in *Biochemistry*, 5th ed., New york: W.H. Freeman, 2002.

[33] B. Kok, B. Forbush, and M. McGloin, “Cooperation of Charges In Photosynthetic O2 Evolution - I. A Linear Four Step Mechanism,” *Photochem. Photobiol.*, vol. 11, pp. 457–475, 1970.

[34] M. R. Razeghifard and R. J. Pace, “Electron Paramagnetic Resonance Kinetic Studies of the S States in spinach PSII membranes,” *Biochim. Biophys. Acta*, vol. 1322, pp. 141–145, 1997.

[35] N. Cox, J. L. Hughes, R. Steffen, P. J. Smith, A. W. Rutherford, R. J. Pace, and E. Krausz, “Identification of the QY Excitation of the Primary Electron Acceptor of Photosystem II: CD Determination of Its Coupling Environment,” *J. Phys. Chem. B*, vol. 113, no. 36, pp. 12364–12374, Sep. 2009.

[36] K. Acharya, B. Neupane, V. Zazubovich, R. T. Sayre, R. Picorel, M. Seibert, and R. Jankowiak, “Site Energies of Active and Inactive Pheophytins in the Reaction Center of Photosystem II from Chlamydomonas reinhardtii,” *J. Phys. Chem. B*, vol. 116, no. 12, pp. 3890–3899, Mar. 2012.

[37] M. Germano, C. C. Gradinaru, A. Y. Shkuropatov, I. H. M. van Stokkum, V. A. Shuvalov, J. P. Dekker, R. van Grondelle, and H. J. van Gorkom, "Energy and Electron Transfer in Photosystem II Reaction Centers with Modified Pheophytin Composition," *Biophys. J.*, vol. 86, no. 3, pp. 1664–1672, Mar. 2004.

[38] L. Zhang, D.-A. Silva, H. Zhang, A. Yue, Y. Yan, and X. Huang, "Dynamic protein conformations preferentially drive energy transfer along the active chain of the photosystem II reaction centre.," *Nat. Commun.*, vol. 5, no. May, p. 4170, 2014.

[39] L. Gagliardi, R. Lindh, and G. Karlström, "Local properties of quantum chemical systems: The LoProp approach," *J. Chem. Phys.*, vol. 121, no. 10, pp. 4494–4500, Sep. 2004.

[40] E. A. Coutsias, C. Seok, and K. A. Dill, "Using quaternions to calculate RMSD," *J. Comput. Chem.*, vol. 25, no. 15, pp. 1849–1857, 2004.

[41] V. I. Novoderezhkin, J. P. Dekker, and R. van Grondelle, "Mixing of Exciton and Charge-Transfer States in Photosystem II Reaction Centers: Modeling of Stark Spectra with Modified Redfield Theory," *Biophys. J.*, vol. 93, no. 4, pp. 1293–1311, Aug. 2007.

[42] K. Riley, R. Jankowiak, M. Rätsep, G. J. Small, and V. Zazubovich, "Evidence for Highly Dispersive Primary Charge Separation Kinetics and Gross Heterogeneity in the Isolated PS II Reaction Center of Green Plants†," *J. Phys. Chem. B*, vol. 108, no. 29, pp. 10346–10356, Jul. 2004.

8. Appendix

8.1. Proteins

A polymer is a large molecule composed of molecular subunits called residues. Proteins are polymers composed of sequences of amino-acid residues bonded together. Amino-acids are composed of two parts, a backbone and a side-chain, also known as an R group. There are twenty different types of amino-acid residues, and each type can be identified by its unique side-chain, while all amino-acid residues share the same backbone structure. Figure 8-1 illustrates a generic amino-acid, illustrating the structure of the backbone composed of four hydrogen atoms (H), one nitrogen atom (N), two carbon atoms (C) and two oxygen (O) and backbone's connection to the R group side-chain. These amino acid residues form polymers as they are "chained" together through their backbones. The "chains" are made through peptide bonds between the carbon bonded to the two oxygen and the nitrogen of the second amino acid residue. The oxygen and the hydrogen bonded together are thusly only present once in an amino acid chain at one of the ends of the chain labeled the C-terminus (also known as the carboxyl-terminus). Similarly, only a single nitrogen atom is bonded to two hydrogen atoms in protein, this amino acid is known as the N-terminus (also known as the amino-terminus). A two chain protein is shown in Figure 8-2.

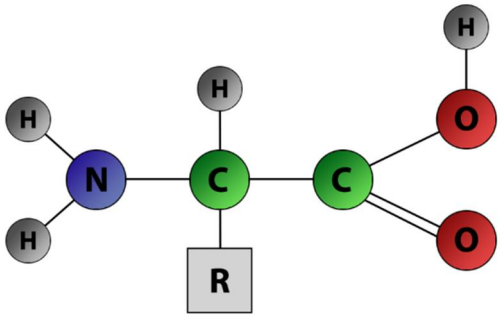

Figure 8-1 - The generic structure of an alpha amino acid in its un-ionized form.

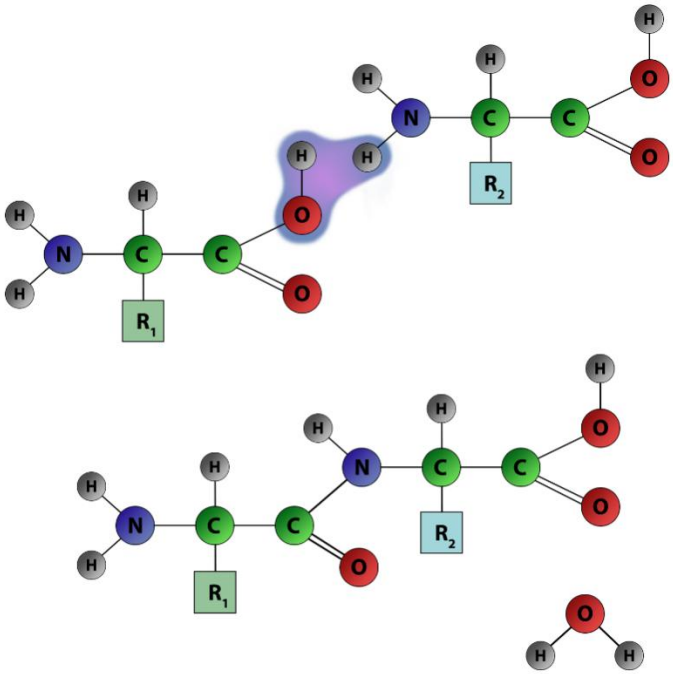

Figure 8-2 - The condensation of two amino acids to form a dipeptide through a peptide bond

8.2. PDB Files and Hydrogen Naming Conventions

The Protein Data Bank (PDB) file format contains molecular structure information typically derived from x-ray diffraction and nuclear magnetic resonance analysis. The molecular information is stored as a list of atoms and the information for each atom typically contains a serial number, name, residue name, chain identifier, residue name, residue sequence number and Cartesian coordinates. There are additional fields (alternate atom location, code for insertion of residues, occupancy, temperature factor, segment identifies, element symbol and charge) which typically do not get used and many programs will not read the information from. PDB files are represented as text files, with each line in the file representing an atom's information (or some other piece of information). The first 6characters specify the information type on the line ("ATOM " indicates the line is an atom record), characters 7 through 11 are the atom's serial number, characters 13 through 16 are for the atom name (atom names are unique amongst all atoms in a residue), character 17 is an alternate location field which links to other atom records to indicate ambiguity in an atom's coordinates from the x-ray analysis, characters 18-20 are for the residue name/type, characters 22 identifies the chain (of residues) in which the atom belongs to, character 27 code for insertion of residues, characters 31 through 54 are the three-dimensional Cartesian coordinates of the atom, characters 55 through 60 are the occupancy of the atom (percentage of times atom is present in crystal structure), characters 61-66 are the temperature factor, characters 73 through 76 are the segment identifier, characters 77 through 78 are for the element symbol and characters 79 through 80 may give the charge of the atom.

Not every character space in the line of 80 characters is used and many fields will generally be omitted from the file. The atom symbol is a field generally omitted and

thus many software packages will determine the atom type from the atom name, which the PDB file format tries to be in agreement with IUPAC (International Union of Pure and Applied Chemistry) naming standards (though has been unable to in cases). Atom names are unique to their residue and their field is composed of four characters (characters 13-16) in the line of atom information. The first two characters are reserved for the atom symbol and for atoms with a single character symbol the first character is a space. As the atom symbol field in PDB files is typically omitted, many software packages will determine the atom type from the atom's name. Characters three and four of the atom name field can provide extra information about the atom, such as bond information. As an example the "first" carbon in a residue might be labeled "Carbon A" and thus its name would be "CA" and entered "_CA_" into the name field in the PDB file. If three hydrogen atoms were bonded to "CA" their names would be "HA1", "HA2" and "HA3". Respectively they would have their names recorded in the PDB file as "_HA1", "_HA2" and "_HA3" indicating that they are hydrogen atoms bonded to atom "A".

Though "Carbon A" ought to be differentiable from "Calcium" in a PDB as their names would appear "_CA_" and "CA___" respectively, there is an issue of handling atoms with single character atom symbols and four character names. This is an issue which is quite common for hydrogen atoms to quite frequently have four character names, and will not have their name fit into the PDB name field while complying with the standards. Different software packages have devised their own solutions to the problem, and unfortunately no standard has emerged, breaking direct compatibility of PDB between different software packages. There are four prominent solutions to the hydrogen naming problem which I have labeled "Spill Over", "Wrap

Around", "Wrap Around Extreme" and "Left Shift" which are described in the following sections. In order to preserve and not pervert information, diligence is required when loading an output PDB file from one software package into another.

8.2.1. "Spill Over"

"Spill Over" is a solution which adheres to the PDB specification that the first two (of four) characters for the atom name are the atom type's symbol and takes advantage that hydrogen atoms do not show up in x-ray structures and thus will never have alternate locations, leaving character 17 (the character field proceeding the atom name) to be guaranteed to be unused for its designated purpose with respect to hydrogens. With this method, the first character of the atom name field is a space and the second field is an "H". The third and fourth character fields of the atom name are the second and third characters of the atom's name. The fourth (last) character of the atom's name is then placed into character field 17. This method is helpful because there is never any ambiguity with regards to the type of atom, however many software packages will not read the entire atom name. For example a hydrogen atom with the name "HA12" would be entered into the name field of the atom record as "_HA1" and the "2" would be entered into the alternate location field.

8.2.2. "Wrap Around"

The "Wrap Around" solution places the hydrogen symbol "H" into the appropriate second column of the name field and the following two characters of the atom's name into columns three and four. The fourth character of the atom's name is placed into the first column of the name field. A hydrogen atom with the name "HA12" would be entered into the name field as "2HA1".

8.2.3. "Wrap Around Extreme"

The “Wrap Around Extreme” implements a similar strategy to that of “Wrap Around” in section 8.2.2 by placing the last character of the hydrogen atom’s name into the first column of the name field, however this method adheres to different rules as to when this is done. The fourth character is moved into the first column of the name field when the hydrogen atom's name consists of four characters or when the atom’s name is three characters long and both characters of proceeding the “H” are numeric. Thus, like “Wrap Around” “HA12” would be entered as “2HA1” however now “H12” would be entered as “2H1_” unlike “Wrap Around” which would enter it “_H12”.

8.2.4. "Left Shift"

The “Left Shift” solution will shift each atom name character over one space to the left, so that the "H" appears in the first column of the name field, the second column contain the second character of the atom’s name, likewise the third and fourth columns contain the third and fourth characters of the atom’s name respectively. The "Left Shift" technique leaves room for ambiguity, as a hydrogen atom with the name "HE**" (where * is any character) will be interpreted as a helium atom by many software packages.

8.3. Amber’s LEaP

Amber is a set of molecular mechanical force fields and a package of molecular simulation programs. These force fields and molecular simulation programs are used for the simulation of biomolecules.

8.4. Quaternions

Quaternions extended the complex number plane and are represented as four dimensional vectors with basis elements customarily denoted $\hat{1}, \hat{\imath}, \hat{\jmath}$ and $\hat{k}$. $\hat{1}$ is the basis element for quaternions, thus quaternion q when written in linear combination form is written

$$q = a + b\hat{\imath} + c\hat{\jmath} + d\hat{k} \tag{8.1}$$

Where a, b, c and d are real numbers. $a + 0\hat{\imath} + 0\hat{\jmath} + 0\hat{k}$ is real, while $0 + b\hat{\imath} + c\hat{\jmath} + d\hat{k}$ is pure imaginary. The products of basis elements is defined as such:

$$\begin{aligned} \hat{\imath}^2 = \hat{\jmath}^2 = \hat{k}^2 = \hat{\imath}\hat{\jmath}\hat{k} = -1 \\ \hat{\imath}\hat{\jmath} = \hat{k}, \qquad \hat{\jmath}\hat{\imath} = -\hat{k} \\ \hat{\jmath}\hat{k} = \hat{\imath}, \qquad \hat{k}\hat{\jmath} = -\hat{\imath} \\ \hat{k}\hat{\imath} = \hat{\jmath}, \qquad \hat{\imath}\hat{k} = -\hat{\jmath} \end{aligned} \tag{8.2}$$

The addition of two quaternions $q =< q_1, q_2, q_3, q_4 >$ and $r =< r_1, r_2, r_3, r_4 >$ is defined as

$$q + r =< q_1 + r_1, q_2 + r_2, q_3 + r_3, q_4 + r_4 > \tag{8.3}$$

The multiplication between q and r is called the Hamilton product and using the distributive law and product of basis defined in equation (8.2) gives the result

$$\begin{aligned} q \cdot r =< & q_1r_1 - q_2r_2 - q_3r_3 - q_4r_4, \\ & q_1r_2 + q_2r_1 + q_3r_4 - q_4r_3, \\ & q_1r_3 - q_2r_4 + q_3r_1 - q_4r_2, \\ & q_1r_4 - q_2r_3 - q_3r_2 - q_4r_1 > \end{aligned} \tag{8.4}$$

and as can be seen in equation (8.4), quaternion multiplication is noncommutative. The conjugate of quaternion q is given by q^* where

$$q^* = a - b\hat{\imath} - c\hat{\jmath} - d\hat{k} \tag{8.5}$$

8.5. Residue Partitions

The following figures illustrate PSII-RC cofactors and how they were separated into their partitions.

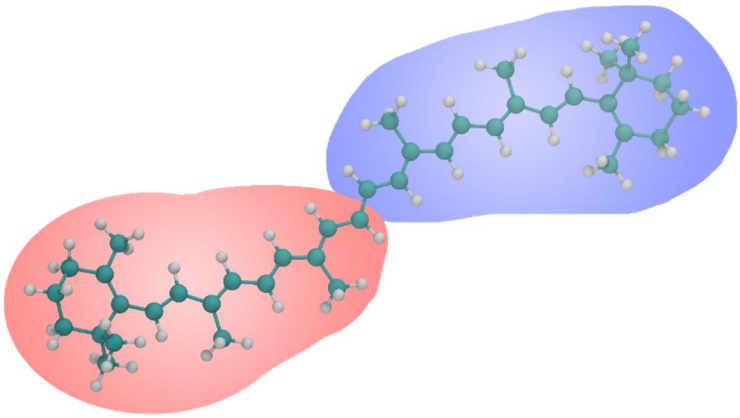

Figure 8-3 - Beta-Carotene (BCR) with its two fragments used for environment selection highlighted. Fragment 1 (BC1) highlighted in red and fragment 2 (BC2) highlighted in blue.

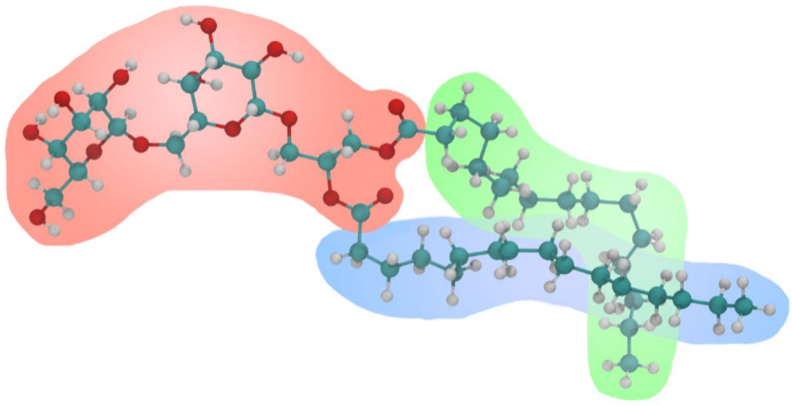

Figure 8-4 – Digalactosyl Diacyl Glycerol (DGD) with its three fragments used for environment selection highlighted. Fragment 1 (DG1) highlighted in red, fragment 2 (DG2) highlighted in blue and fragment 3 (DG3) highlighted in green.

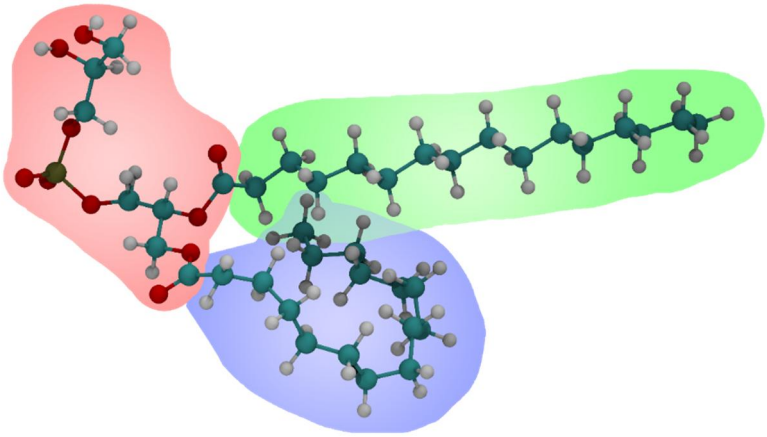

Figure 8-5 - Dipalmitoyl-Phosphatidyl-Glycerole (LHG) with its fragments highlighted. Fragment 1 (LH1) highlighted in red, fragment 2 (LH2) highlighted in blue, fragment 3 (LH3) highlighted in green.

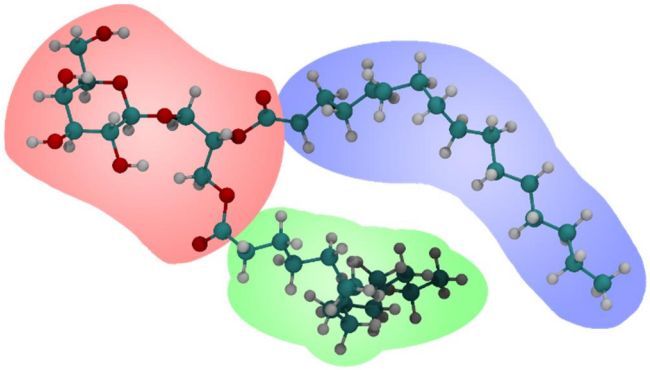

Figure 8-6 - Disetearoyl-Monogalactosyl-Diglyceride (LMG) with its fragments highlighted. Fragment 1 (LM1) highlighted in red, fragment 2 (LM2) highlighted in blue, fragment 3 (LM3) highlighted in green.

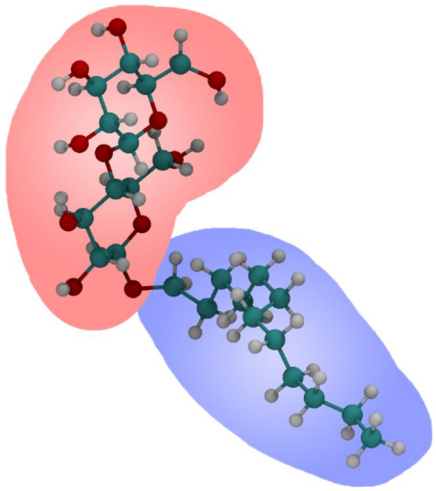

Figure 8-7 - Dodecyl-Beta-D-Maltoside (LMT) with its fragments highlighted. Fragment 1 (LM1) highlighted in red, fragment 2 (LM2) highlighted in blue.

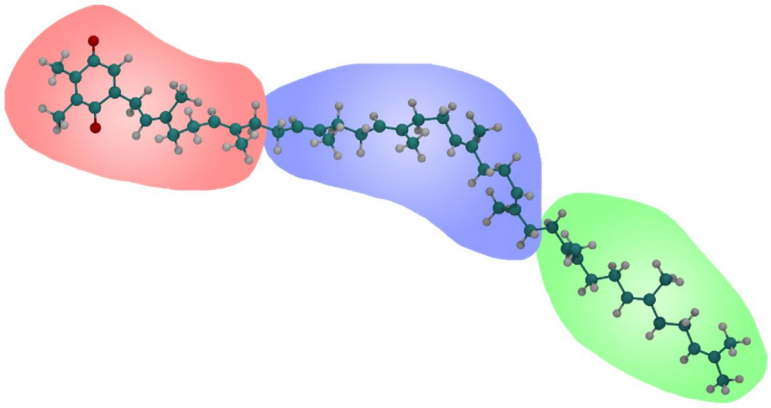

Figure 8-8 - Plastoquinone 9 (PL9) with its fragments highlighted. Fragment 1 (PL1) highlighted in red, fragment 2 (PL2) highlighted in blue, fragment 3 (PL3) highlighted in green.

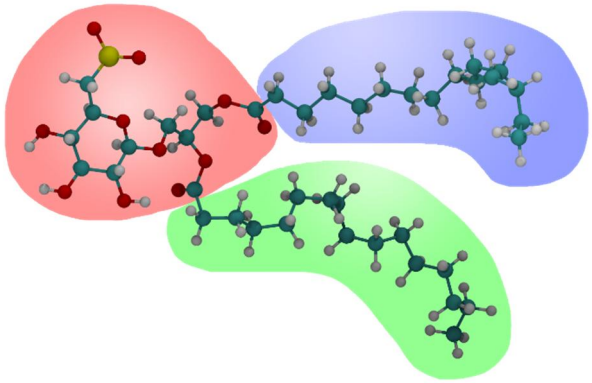

Figure 8-9 - Sulfoquinovosyldiacylglycerol (SQD) with its fragments highlighted. Fragment 1 (SQ1) highlighted in red, fragment 2 (SQ2) highlighted in blue, fragment 3 (SQ3) highlighted in green.

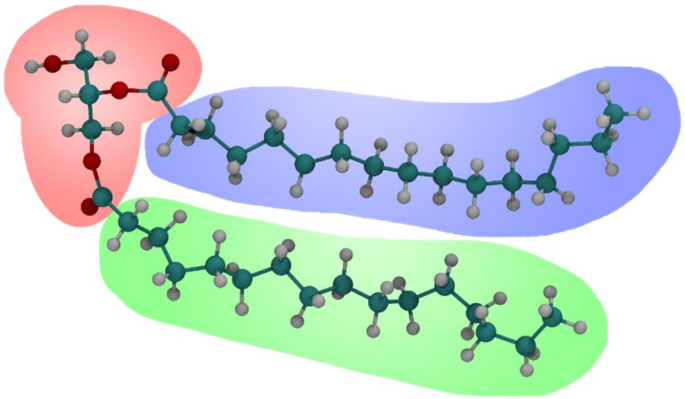

Figure 8-10 - Hydroxytetradecanoyl (UDG) with its fragments highlighted. Fragment 1 (UD1) highlighted in red, fragment 2 (UD2) highlighted in blue, fragment 3 (UD3) highlighted in green.

8.6. Developed Software

8.6.1. Helper Classes

Atom.h / Atom.cpp – Contains the class Atom, which holds all of the atom properties found in an ATOM record in a PDB file in addition to dipole, quadrupole and polarizability information as well as exclusion lists for DPFs. Additionally there are functions for the Atom class to retrieve the atom's symbol from its name, compute the distance between the atom and another atom, and check whether the atom belongs to an amino acid and to get the atom in either PDB or XYZ format.

Library.h / Library.cpp – The library class is an extension of the **Potential** class used for creating and loading Library files. Library files are similar to Dalton Potential files with the addition of assigning atoms into groups. Groups are used in fitting atom (and bond center) properties of from one structure to another as described in section 4.2.2. When fitting one structure to another the residues are separated into their groups and each group is then fitted individually.

Menu.h / Menu.cpp - A complimentary class, which processes command line inputs placing the arguments into their respective variables. This class will also ensure that all required fields have been provided in the arguments and checks if help has been requested ("-h" or "-H"), setting respective flags if either case holds true.

PDB.h / PDB.cpp – The PDB class stores a collection of Atoms ("Atom.h") in a std vector with structures of type "Chain", which hold std vectors of type Residues, which hold std vectors of type Atom. There are iterators available for the PDB class as well as Chain and Residue structures which iterate over all atoms scoped within their respective containers in order that the atoms would appear in the PDB file. The PDB class has helper functions to both load and save a PDB file allowing for

any of the hydrogen naming conventions to be used which are described in section 8.2.

Potential.h / Potential.cpp – A class which holds information from Dalton Potential Files (DPF). The information is stored in a std vector of type Atom ("Atom.h") and provides functions to load a DPF, MOL2 file or XYZ file. There are also functions for saving the exclusion lists as DPF or XYZ (for visualisation in VMD).

8.6.2. Programs for ONIOM Optimization

1. **partition_residues.cpp** – Takes an input PDB file and a list of partition definition files, separates all residues with partition definitions into their partitions. It then finds residues which will compose the cores in the ONIOM optimization and selects all residue partitions within a cut-off distance of the cores. A PDB file is saved for each of the cores which include the cores (which are marked using the tempfactor column of the PDB file) and their local environments.
2. **CapAcidsWithLeap.sh** – This script uses LEaP to add caps to uncapped residues in a PDB file. Its typical use case is to cap residues which have had bonds broken by **partition_residues.**
3. **CreateG09ONIOMInputFile.cpp** – This program converts a PDB file (generally processed by **CapAcidsWithLeap**) into an ONIOM input file for Gaussian. It sets the cores as high level and their environments at low level. Additionally it removes hydrogen caps from residue partitions whose neighbouring residue partitions resides within the file. It also caps amino acids not

in a chain, as LEaP does not have definitions for these and is thus unable to cap them.

8.6.3. Programs for Site Energy Computations

1. **seperatePDB.cpp** – Takes a PDB file and for each residue in the file saves a PDB file containing only that single residue. The PDB files generated are used by MolCas to compute polarizabilities and multipoles.
2. **mol2TOdalton.cpp** – Converts a MolCas output file which contains polarizabilities and multipoles of atoms and bond centers) into a Dalton Potential File (DPF).
3. **MergePotentials.cpp** –Uses multiple DPFs as input and combines them into a single DPF as described in section 4.2.1. Additionally the combined DPFs maybe be outputted as a PDB file or as an XYZ file for structural and exclusion list visualisation respectively. As DPF files do not include bond information, atom bonds are determined by finding the closest two equidistant atoms to each bond center. Optionally the number of bonds to grow exclusion lists into neighbouring amino acids can be configured and the program can warn be told to throw warnings if a residue's charge is not within a certain distance of a natural number.
4. **SelectPotentials.cpp** – Used to clear atoms and bond centers of properties (ie. Multipoles, polarizabilities) not within a cut off distance of a given structure. The programs loads a DPF and a XYZ file and removes specified properties from atoms and bond centers who do not sit within a given proximity of any of the atoms of the XYZ file.
5. **create_lib_file.cpp** – Create a Library file (.lib) from either a MolCas output file or a PDB file and a DPF. The program will automatically assign atoms

and bond centers to groups. The outputted Library file is used for fitting by **rotate_groups**.

6. **rotate_groups.cpp** – Fits a Library (a list of library files (.lib)) containing atom and bond center properties of various residues and their groups) to the structure of a PDB file. The fitting is performed as described in section 4.2.2. The output is a DPF (and optionally a PDB for visualisation) with the atom coordinates from the provided PDB and the atom properties of the Library files fitted to that PDB. There are parameters which can be modified such as the acceptable RMSD range when fitting (default 0.1) and the number of atom bonds to extend amino acid exclusion lists into neighbouring amino acids. Optionally the exclusion lists may be saved as a XYZ file which will allow visualisation of the exclusion lists in software such as VMD.

output_charge.cpp – Loads either a Potential file or a Library file and outputs the total charge of all of the atoms.

remove_atoms_from_library.cpp – Loads a Library file and removes all atoms with specified indices and/or names, saving results into a Library file.

save_exclusion_lists.cpp – Used for visualising exclusion lists. Takes a DPF and saves an XYZ file with exclusion lists. Used to create files to visualise exclusion lists in software such as VMD.